Investigation of Proton Transfer Pathways in Bacteriorhodopsin with Multi-Length-Scale Simulations

Von der Fakultät für Lebenswissenschaften

der Technischen Universität Carolo-Wilhelmina

zu Braunschweig

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

genehmigte

D i s s e r t a t i o n

von Prasad Vishwas Phatak

aus Mumbai /Indien

1. Referent: Professor Dr. Marcus Elstner
2. Referent: Professor Dr. Karl-Heinz Gericke
Eingereicht am: 06.04.2009
Mündliche Prüfung (Disputation) am: 08.07.2009
Druckjahr 2009

Vorveröffentlichungen der Dissertation

Teilergebnisse aus dieser Arbeit wurden mit Genehmigung der Fakultät für Lebenswissenschaften, vertreten durch den Mentor der Arbeit, in folgenden Beiträgen vorab veröffentlicht:

1. *Amino acids with an intermolecularpProton bond as proton storage site in bacteriorhodopsin*
   Phatak*, P.; Ghosh*, N.; Yu. H.; Cui, Q. and Elstner, M. *Proc. Natl. Acad. Sci. USA* **2008**, 105, 19672-19677.
   * These authors contributed equally to this work.

2. *Long-distance proton transfer with a break in the bacteriorhodopsin active site*
   Phatak, P.; Frähmcke, J.S.; Wanko, M.; Hoffmann, M.; Strodel, P.; Smith, J.C.; Suhai, S.; Bondar, A.N. and Elstner, M. *J. Am. Chem. Soc.* **2009**, 131, 7064-7078.

Parts of this thesis were published in (1) and (2).

Abstract

All living organisms need energy to carry out diverse biochemical reactions. Processes such as transformation, storage and utilization of energy are common in all living organisms irrespective of their nature. The storage of available energy into chemically useful form involves complex biochemical reactions. These reactions are usually coupled to the generation of a proton-motive electrochemical gradient across a membrane through proton-pumping proteins.

A typical example of a light-activated ion-pumping protein is bacteriorhodopsin, which is found in the cell membrane of a halophilic archaeon *Halobacterium salinarum*. In bacteriorhodopsin, one proton is transferred across the cell membrane through a photocycle that consists of five conformation-coupled proton transfer (PT) steps. Despite the knowledge gained from numerous crystallographic and spectral studies, several issues remain unclear that are essential for an understanding of the PT mechanism in bacteriorhodopsin.

Recent advances in computer simulation techniques have extended their applicability to study complex biological processes. In case of bacteriorhodopsin, numerous previous and ongoing simulation studies have been focussed on the events in the early half of the photocycle. The later half of the photocycle events are challenging to investigate due to involved complex and long-range PT reactions and significant contributions from entropic and solvent effects. In the current work, hybrid quantum mechanics/molecular mechanics simulations were carried out to investigate three unsolved problems associated with the late photocycle events which take place on the extracellular side of bacteriorhodopsin.

The identity of the proton release group (PRG) located at the extracellular side of bacteriorhodopsin has been controversial. Over the years, several protein groups were proposed to function as the PRG. In particular, FTIR as well as simulation studies indicated the PRG to be formed by a protonated water cluster stabilized by surrounding protein groups including Glu194 and Glu204, with a characteristic continuum band in the 1800-2000 cm^{-1} region. However, none of the proposed models can consistently explain all experimental and theoretical results. The simulations carried out in this work show that a proton partially delocalized between Glu194 and Glu204 residues gives rise to a continuum band similar to that observed in the FTIR studies. The proposed model of the PRG that consists of a

proton shared between Glu194 and Glu204 is also consistent with the structural models suggested by X-ray crystallography.

In bacteriorhodopsin, the last photocycle step occurs during O→bR transition and involves the transfer of a proton from Asp85 to the PRG. In comparison to the early half of the photocycle, much less is known about the O state and the PT mechanism. It is also unclear whether the PT involves any transient photo-intermediate(s) as suggested by many experiments. Extensive minimum energy path calculations augmented with UV-Vis and vibrational normal mode analysis were carried out to investigate the structural details of the O state. The simulations suggest that the retinal is twisted in the O state and atleast three water molecules are present in the active site of the O state. The results also indicate that O→bR transition could indeed involve a transient O state characterized by deprotonated Asp85 and protonated Asp212.

To investigate the long-range PT that takes place during O→bR transition, minimum energy path calculations were carried out in this work. Different O-like models were studied to determine the orientation of Arg82 and to assess its role in the PT. The calculations indicate that Arg82 is more likely to adopt an extracellular orientation in the O state which can enable the formation of a chain of hydrogen bonded water molecules in the active site. The excess proton may then be transferred from Asp85 *via* Asp212 to the PRG.

Zusammenfassung

Alle lebenden Organismen brauchen Energie, um diverse biochemische Reaktionen durchzuführen. Prozesse wie Umwandlung, Speicherung und Nutzung von Energie sind allen lebenden Organismen artunabhängig gemeinsam. Die Speicherung von verfügbarer Energie in einer chemisch nutzbaren Form erfordert komplexe biochemische Reaktionen. Diese Reaktionen sind oft gekoppelt an die Erzeugung eines durch Protonen verursachten elektrochemischen Gradienten an Membranen durch protonenpumpende Proteine.

Ein typisches Beispiel eines lichtgetriebenen ionenpumpenden Proteins ist Bakteriorhodopsin, das in der Zellmembran des halophilen Archaeons *Halobacterium salinarum* gefunden wird. In Bakteriorhodopsin wird ein Proton in einem Photozyklus, der aus fünf konformationsgekoppelten Protonentranferen (PT) besteht, durch die Zellmembran geleitet. Trotz der Erkenntnisse aus mehreren kristallographischen und spektroskopischen Studien sind einige Fragestellungen unklar, die für den Mechanismus des Protonentransfers in Bakteriorhodopsin essentiell sind.

Neueste Fortschritte in Computersimulationstechniken haben deren Anwendbarkeit im Studium komplexer biologischer Prozesse ausgeweitet. Im Falle von Bakteriorhodopsin haben sich viele abgeschlossene und laufende Simulationsstudien auf die Ereignisse in der frühen Hälfte des Photozyklus konzentriert. Die Untersuchung der Vorgänge in der späten Hälfte des Photozyklus ist anspruchsvoll, weil komplexe und weitreichende PT-Reaktionen und signifikante Beiträge von Entropie und Solvatationseffekten eine wichtige Rolle spielen. In der vorliegenden Arbeit wurden gemischte quantenmechanische / molekularmechanische Simulationen durchgeführt, um drei ungelöste Probleme zu untersuchen, die mit den Ereignissen an der außerzellulären Seite von Bakteriorhodopsin in der späten Hälfte des Photozyklus zusammenhängen.

Die Zusammensetzung der Protonenabgabegruppe (PRG), die an der außerzellulären Seite von Bakteriorhodopsin liegt, ist umstritten. FTIR sowie Simulationen deuten an, dass die PRG durch einen protonierten Wassercluster gebildet wird, der von den umgebenden Proteingruppen, besonders Glu194 und Glu204, stabilisiert wird. Dieser hat eine charakteristische Kontinuumbande in der 1800-2000 cm^{-1} Region. Aber keines der vorgeschlagenen Modelle kann alle experimentellen und theoretischen Resultate konsistent erklären. Die in dieser Arbeit

durchgeführten Simulationen zeigen, dass ein partiell delokalisiertes Proton zwischen den Gruppen Glu194 und Glu204 der Grund für eine Kontinuumbande sein kann ähnlich der in FTIR Studien beobachteten. Das beschriebene Modell des geteilten Protons zwischen Glu194 und Glu204 ist auch konsistent mit dem von der Röntgenstrukturanalyse vorgeschlagenen Modell.

Der letzte Protonentransferschritt in Bakteriorhodopsin findet im O→bR Übergang statt und beinhaltet den Transfer eines Protons von Asp85 zur PRG. Im Vergleich zu der frühen Hälfte des Photozyklus ist über den O-Zustand und den PT-Mechanismus viel weniger bekannt. Desweiteres ist unklar, ob der PT ein oder mehrere kurzlebige Photointermediate enthält, die experimentell vorgeschlagen wurden. Umfangreiche Minimal-Energie-Pfad-Berechnungen unterstützt von UV-Vis und Normalschwingungsmoden-Analyse wurden durchgeführt, um strukturelle Details des O-Zustandes aufzuklären. Die Simulationen deuten darauf hin, dass das Retinal im O-Zustand getwistet ist, und sich mindestens drei Wassermoleküle in der aktiven Region befinden. Die Ergebnisse weisen weiter darauf hin, dass der O→bR Übergang in der Tat einen kurzlebigen O-Zustand beinhalten kann, der durch ein deprotoniertes Asp85 und ein protoniertes Asp212 charakterisiert ist.

Um den weitreichenden PT zu untersuchen, der während des O→bR Übergangs stattfindet, wurden in dieser Arbeit Minimal-Energie-Pfad-Berechnungen durchgeführt. Verschiedene O-ähnliche Modelle wurden untersucht, um die Orientierung des Arg82 und seine Rolle im PT zu bestimmen. Die Rechnungen deuten darauf hin, dass Arg82 im O-Zustand wahrscheinlich eine außerzelluläre Position einnimmt, die die Bildung eines Kanals von verbrückten Wassermolekülen zur aktiven Region ermöglichen kann. Das überschüssige Proton kann dann von Asp85 *via* Asp212 zur PRG transferiert werden.

Contents

List of Figures

List of Tables

Chapter 0

Prologue

The life on earth consists of a large diverse array of organisms ranging from micro-scale prokaryotic species such as archea and bacteria, to large-scale eukaryotic species such as plants and animals. Any living organism, irrespective of size and shape can be differentiated from a non-living object by its ability to carry out processes such as growth, metabolism, and response to various stimuli. Numerous chemical reactions that occur in an organism as a part of its metabolism, include production and utilization of energy which is stored in adenosine triphosphate (ATP). The synthesis of ATP by an enzyme, ATP synthase, involves harnessing energy produced from the motion of protons under transmembrane electrochemical gradient generated by proton pumps either by using light energy or some form of electrochemical energy. Thus, understanding the mechanism of proton pumps is crucial for understanding how bioenergetic processes occur.

Bacteriorhodopsin is the simplest known light-activated proton pump and was discovered in the cell membrane of an archaeon *Halobacterium salinarum* in early 1970s. Over the years, bacteriorhodopsin has become a paradigm for studying the structure of trans-membrane proteins and for understanding the proton transfer (PT) mechanisms in ionic pumps. In bacteriorhodopsin, effectively one proton is translocated from the cytoplasmic side to the extracellular side of the membrane, through a multi-step reaction process known as the photocycle. The photocycle is initiated after light absorption by the retinal chromophore which undergoes isomerization from all-*trans* to 13-*cis* configuration and involves formation of several intermediate states, characterized by their absorption maxima.

Since its discovery, three dimensional structures of various bacteriorhodopsin

intermediates have been solved initially by electron microscopy and then by X-ray crystallography at increasingly high resolutions. The crystal structures have revealed locations of several crucial amino acid residues and water molecules, and provided hints to the structural changes that take place during the photocycle. Despite available high resolution crystal structures, the lack of positions of protons in the resolved structures inhibit the coherent understanding of several key events that take place during the photocycle.

On the other hand, several other experimental studies of bacteriorhodopsin employed a variety of spectroscopic techniques such as UV-Vis, resonance Raman, Fourier transform infrared (FTIR), and nuclear magnetic resonance (NMR), augmented with techniques such as site-directed mutagenesis and pH titration experiments. In particular, these studies monitored: (i) changes in the spectral fingerprint modes (FTIR and Raman); (ii) variation in the inter-atomic distances, bond and dihedral angles (solid state NMR), and (iii) changes in the absorption maximum of photo-intermediates along the photocycle (UV-Vis).

Combination of results from the crystallographic and spectroscopic experiments have enhanced our understanding of the PT mechanism at molecular level. Nevertheless, currently these techniques can not monitor many dynamical changes that take place in the protein such as details of PT pathways or structural changes associated with protonation or deprotonation of a particular amino acid during the photocycle. Thus albeit our improved understanding of the PT mechanism in bacteriorhodopsin and function of the ion transporting proteins in general; several detailed key issues remain unclear. Further, many of the experimental studies provide conflicting results and thus amplify the gap between present knowledge of the overall process and the coherent understanding of ion transport mechanisms in general.

Development of computers during the 20th century added a new dimension in the understanding of biophysical/biochemical processes. Although computers were employed to solve many complex equations as early as in 1950s, the development of faster computers along with the development of numerical simulation methodologies provide a promising tool that aim to reduce the above mentioned knowledge gap. By allowing us to monitor the dynamical changes at an atomic level, these tools can relate available experimental results to a clear understanding of the protein function and biochemical processes. Furthermore, the computer simulations have provided us with a way to predict events that take place during the

biochemical processes, and thus to predict the rules that govern these processes.

Indeed in recent years, a variety of computer simulation techniques have been used to reveal many interesting details of the bacteriorhodopsin photocycle. Insights gained from these simulations have provided a few of the missing pieces in the jigsaw puzzle of the bacteriorhodopsin PT mechanism. In this doctoral work, this interesting tool of biomolecular computer simulations has been employed to investigate events that take place at the extracellular side of bacteriorhodopsin during the later half of the photocycle. In the following part of the prologue, the details of this work are briefly outlined. The thesis is divided in following three sections.

(i) The first section summarizes available knowledge on bacteriorhodopsin. In the first chapter, the bacteriorhodopsin photocycle is briefly explained. The second chapter summarizes the results from crystallographic and spectroscopic experimental studies. The third chapter describes the findings elucidated from the computer simulation techniques.

(ii) In the second section, a brief introduction of the theoretical methods that form the basis of computer simulations is presented. The fourth chapter outlines the fundamentals of the molecular orbital (MO) theory followed by a summary of developments in *ab initio* and semi-empirical implementations of MO theory, and density functional theory. A brief summary of force field methods is described at the end of the fourth chapter. The fifth chapter introduces multi-length-scale approaches that are used in this work such as hybrid quantum mechanics/molecular mechanics (QM/MM) techniques and an implicit solvent model, generalized solvent boundary potential (GSBP).

(iii) The third section presents the results of this doctoral work. In the sixth chapter, the results of QM/MM-GSBP simulations to investigate the identity of the proton release group (PRG) are presented. The results of structural and spectral investigations of the O state, and the calculations that suggest a possible transient O intermediate are summarized in the seventh chapter. In the eighth chapter, the results of the preliminary calculations of the long-range PT from O$\rightarrow$bR transition are presented. Finally, the summary of the thesis is presented in chapter nine followed by an outline of the scope of future work.

Chapter 1

Introduction to Bacteriorhodopsin

Summary

This chapter summarizes the concepts of bioenergetics and introduces a proton transporting membrane protein, bacteriorhodopsin. Section 1.1 summarizes how various organisms carry out biochemical processes to transform energy into biologically useful form by employing ion transporting proteins such as bacteriorhodopsin. In section 1.2, starting with the discovery of bacteriorhodopsin, its structural organization in the membrane is described. In bacteriorhodopsin, the PT process takes place through its photocycle and it is outlined at the end of section 1.2.

1.1 Bioenergetics

The cell is one of the most basic units of life that is capable of handling energy. Several types of cells are found in living organisms and have unique functions and features. However, a feature common to all living cells is the ability to derive energy from the environment, convert it into biologically useful form, and employ it to carry out various energy-driven processes. Green plant cells, certain bacteria and algae use chlorophyll to extract energy from the captured sunlight and utilize it to convert carbon dioxide and water into organic substances through photosynthesis. All animal cells and many bacteria on the other hand, rely on available nutrients for their existence. Amongst these nutrients, the organic compounds are oxidized by atmospheric oxygen to carbon dioxide and water through cell respiration.

Photosynthesis and cell respiration both involve harvesting energy to form a compound, adenosine triphosphate (ATP) discovered in 1929,[1,2] which can be regarded as the energy currency of all living cells. The synthesis of ATP from adenosine diphosphate and inorganic phosphate utilizes the energy that is liberated during a series of oxidation-reduction (or electron-transfer) reactions that take place during both respiration and photosynthesis. The mechanism by which the electron transport is associated with the ATP synthesis was provided by Peter Mitchell[3] and known as the chemiosmotic theory for which he was awarded the Nobel prize in chemistry in 1978. The basic idea of this theory is that the enzymes of the electron-transport and ATP synthesizing systems are localized in the membrane with a well-defined orientation and are functionally linked to a vectorial transfer of positively charged hydrogen ions, or protons, across the membrane. Thus the electron transport generates an electrochemical proton gradient across the membrane which can serve as a driving force for the ATP synthesis. This electrochemical gradient consists of two components: one due to the hydrogen ion concentration across the membrane and second due to the difference in the electrical potential between the two aqueous phases separated by the membrane. Together these two components form the protonmotive force.

The process of ATP synthesis that takes place during photosynthesis and cell respiration is known as photophosphorylation and oxidative phosphorylation respectively. In eucaryotes, these processes occur inside the cell in specific membrane-enclosed organelles known as chloroplasts and mitochondria respectively. On the other hand, in bacteria, both these processes are associated with the cell membrane. Further, the organisms that belong to the class of archaea show a peculiar capability of generating both light-induced or a respiration-induced electrochemical proton gradient necessary for the ATP synthesis.[4] Archaea, recognized as a third domain of life,[5] are widely studied organisms because of their ability to flourish in extreme environments such as near rift vents in the deep sea at high temperatures and in extremely alkaline or acid waters. A prokaryotic Halobacteria that grow under extreme salt concentrations like in natural salt lakes or dead sea belong to this category.

One such halophilic archaea *Halobacterium salinarum* contains four archaeal rhodopsins in its cell membrane: bacteriorhodopsin, Halorhodopsin (HR) and two sensory rhodopsins (SRI and SRII).[6] These four rhodopsins make up all retinyli-

dene proteins present in the organism according to its genome sequence.[7] Bacteriorhodopsin converts light energy into an electrochemical proton gradient which is used for synthesis of ATP by ATP synthase. Halorhodopsin works as a chloride pump and maintains the iso-osmolarity of the cytoplasm during the cell growth. The remaining two photo-receptor rhodopsins (SRI and SRII) are structurally similar to the ion pumps bacteriorhodopsin and HR but differ in their functionality. SRI is an photo-attractant receptor and induces migration of the archaea towards the orange light and away from harmful UV radiations, while SRII is a blue light-activated repellent receptor that detects the sunlight and guides the cell towards the darkness.[8,9]

1.2 Bacteriorhodopsin

1.2.1 The Purple Membrane

In 1967, Stoeckenius first observed that the cell membrane of an extremely halophilic archaea disintegrates into fragments of different sizes when the salt concentration is removed.[10] These fragments were later isolated and included a characteristic deep purple color fragment called the purple membrane.[11] In 1971, it was discovered that the purple color of the membrane is because of retinal (vitamin A aldehyde) bound to an opsin-like protein called bacteriorhodopsin of molecular weight of 26 kDa [1] with proposed function of a photo-receptor.[12,13] The function was later confirmed with observation of the protein working as a light driven proton pump,[14] thereby utilizing the generated proton gradient for the synthesis of ATP.[15–18] Specific interactions between the retinal chromophore and the bacterio-opsin protein are important[19] for high thermal stability of the purple membrane.[20]

In the purple membrane, bacteriorhodopsin is organized as trimer units (Figure 1.1) forming a two-dimensional hexagonal lattice.[13,22,23] The first structural model of bacteriorhodopsin appeared in 1975.[24] This low resolution map obtained by electron microscopy of titled 2-D samples indicated that the protein in the membrane consists of seven α-helices oriented almost perpendicularly to the membrane surface.[24] This remarkable discovery made bacteriorhodopsin, the first integral membrane protein from which any structural information was derived. Following this structural knowledge, the amino acid sequence of the protein consisting of 248

[1] Dalton is one-twelfth the weight of an atom of ^{12}C after chemist John Dalton.

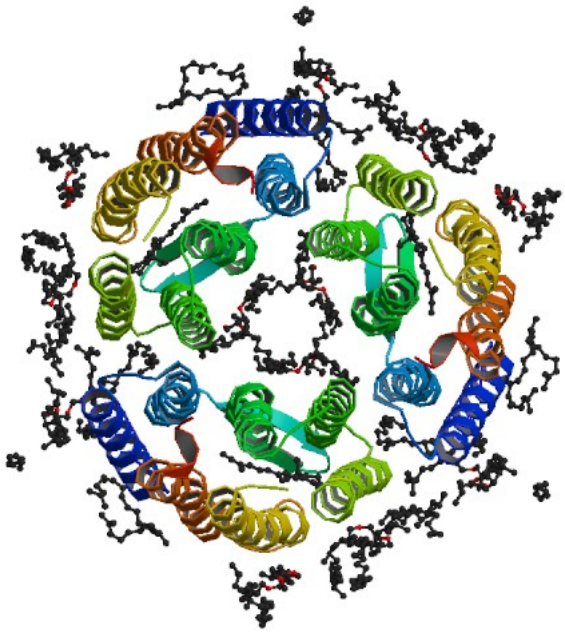


Figure 1.1: The purple membrane of *Halobacterium salinarum* contrains trimeric units of bacteriorhodopsin proteins. The picture prepared from bR structure (Protein Data Bank (PDB) code: 1C3W).[21]

residues was determined.[25–27] Subsequently, the seven helices (labeled from A to G) were assigned to the structural map of bacteriorhodopsin.[28,29]

The lipid molecules account for about 25% by weight of the purple membrane while the remaining is due to the protein.[12,30] The functional significance and location of lipids in the purple membrane has been investigated using several techniques like temperature dependent Fourier transform infrared (FTIR) spectroscopy,[31] neutron diffraction,[32] and absorption spectroscopy[33] (for reviews see refs 34,35). Each hexagonal lattice cell consists of three bacteriorhodopsin monomeric units,[13] and a defined set of glyco- and phospholipid species. Approximately 10 lipid molecules are associated with each bacteriorhodopsin monomeric unit,[32,36] and are located between the monomers and the intermediate space. These molecules are: 2-3 molecules of phosphatidylglycerophosphate methyl ester (PGP-Me), 3 of glycolipid sulphate (S-TGD-1), 1 of phosphatidylglycerol, 1 of Archaea glycocardiolipin (GlyC), 2 of squalene and phosphatidylglycerosulfate (PGS), bispho-

phatidylglycerol (archaeal cardiolipin) (BPG) in small amounts and vitamin MK8 in very small amount.[30,37–39] Later, two novel lipids were discovered in genetically engineered strain of *Halobacterium salinarum*: a phosphosulfoglycolipid (GlyC) and a glycerol diether analog of bisphosphatidylglycerol (BPG).[36] Although these cardiolipin may play an important role in stability of the purple membrane,[40] their location in the purple membrane remains unknown.

The importance of lipid-bacteriorhodopsin interaction was demonstrated by measuring the neutral detergent exposed bacteriorhodopsin trimer using circular dichroism exciton coupling.[41] The lipids are crucial for the stability of the purple membrane,[31] and also for the protein flexibility by maintaining the water hydration shell.[42,43] In addition to this, specific lipids may also be involved in specific functions. For instance, lipid molecules PGP-Me and squalene together are necessary for the normal photocycle behavior of the protein,[33] while S-TGA-1 in combination with PGP-Me are involved in the energy producing ion conductance pathway.[44–46]

In addition to lipids, cations are also bound to bacteriorhodopsin. The influence of removing cations from bacteriorhodopsin and adding native (Mg^{2+} and Ca^{2+}) or other (non-native) cations to deionized bacteriorhodopsin has long been studied. It was observed that removal of native cations blue shifts the absorption maximum by $\sim$0.1 eV[47–49] and ceases proton pumping activity, while addition of native or non-native cations to deionized bacteriorhodopsin (blue membrane) regains the purple membrane and restores the pumping activity.[50–52] However, location of the cations as well as the mechanism by which they influence the photocycle kinetics remains subtle. Further, X-ray crystal structures did not reveal their location(s) either due to unknown effect(s) of crystallization procedure on these cations or due to lack of electron densities of the bound cations in the crystal structures.

Nevertheless, to explain the effect of cation binding on bacteriorhodopsin, two models have been proposed. First is the specific binding model which suggests specific binding site(s) of the cations to the protein possibly on the extracellular side of the protein and/or near retinal pocket on the other side the protein.[50–57] The second model suggests non-specific, random cations (and protons) binding to the negatively charge membrane and thus affects the surface pH according to the Gouy-Chapman theory and influences the purple-to-blue transition.[49,58–61]

The retinal geometry consists of a conjugated polyene chain with a β-ionone ring and has two carbon atoms C_9 and C_{13} with methyl groups instead of hydrogen

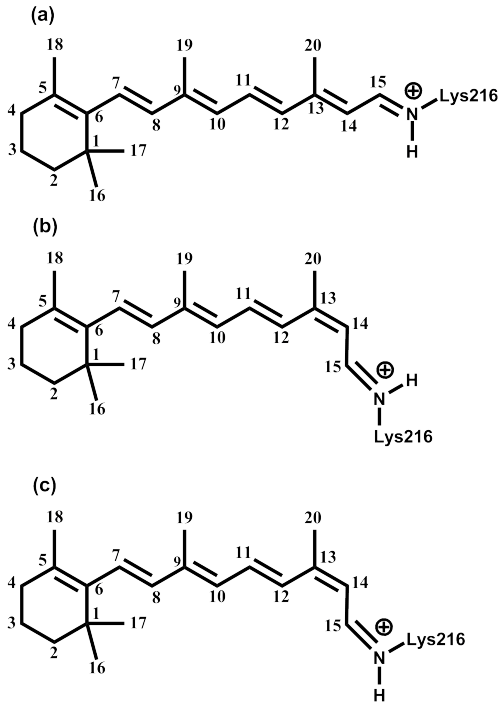


Figure 1.2: Different configurations of the retinal chromophore. (a) all-*trans*; (b) 13-*cis*, 15-*anti*; and (c) 13-*cis*, 15-*syn* configuration. The retinal is covalently bound to Lys216 side-chain *via* a protonated Schiff base.

atoms (Figure 1.2). The binding site of the retinal attached as a protonated Schiff base[62] to the ε-amino group of a lysine residue[12,63] that was initially suggested to be Lys41[26,64] and was later revised to Lys216 on helix G.[65,66] The retinal separates the protein into a cytoplasmic and an extracellular part. The average retinal plane is oriented nearly perpendicular to the plane of the purple membrane[67] and its β-ionone ring side is titled towards the extracellular side of the protein.[68] In the dark, the purple membrane contains a mixture of all-*trans* (bR_{568}) and 13-*cis*, 15-*syn* (bR_{548}) conformers of retinal (Figure 1.2) in 2:1 ratio.[69,70] Under physiological conditions, only the protein with all-*trans* retinal is biologically active (pumps protons) while the protein with 13-*cis* retinal does not pump protons,[71] but instead has a branching pathway such that it converts to the protein with all-*trans* retinal via a process known as light adaptation.[72,73]

The retinal is surrounded by three tryptophane (Trp86, Trp182, and Trp189)

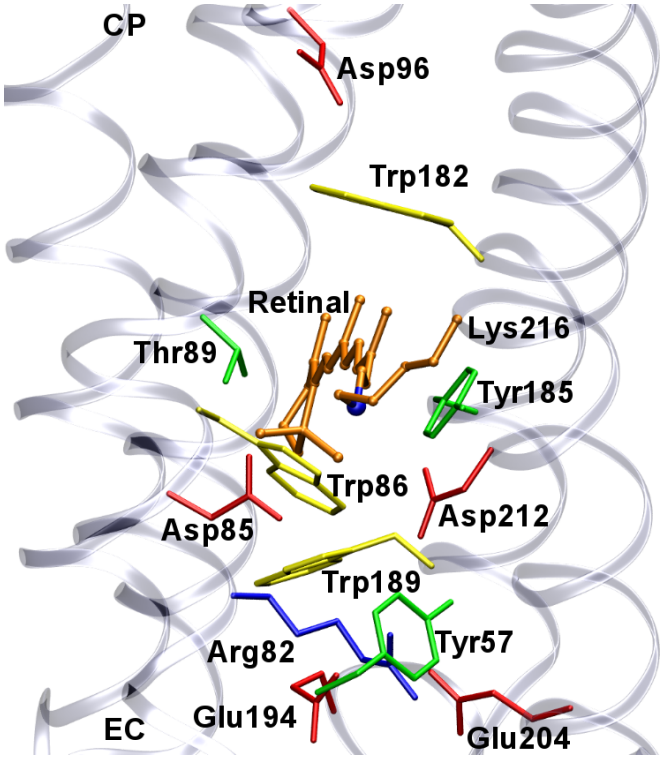


Figure 1.3: A snapshot of bacteriorhodopsin structure indicating location of various important amino acid residues. CP and EC indicate the cytoplasmic and extracellular side respectively. Color codes of various protein groups are: nonpolar residues (Trp86, Trp182 and Trp189) in yellow; polar residues (Tyr57, Thr89, and Tyr185) in green, acidic residues (Asp85, Asp96, Glu194, Asp212, Glu204) in red and basic residues (Arg82) in blue and the retinal is represented in orange. The picture is prepared from bR structure of ref 21 using VMD package.[74]

and two tyrosine residues (Tyr57 and Tyr185) which tightly pack the retinal (Figure 1.3) and provide structural constraints such that the retinal polyene chain is in proper orientation.[75] The positive charge of the protonated Schiff base is stabilized by the counterion residues Asp85 and Asp212. In addition to these, there are several water molecules in the retinal binding pocket,[21,76] which are together known as the hydrogen bonded network (HBN). This HBN is connected to the extracellular proton release group (PRG) *via* Tyr57, positively charged Arg82 and other extracellular water molecules.

1.2.2 Bacteriorhodopsin Photocycle

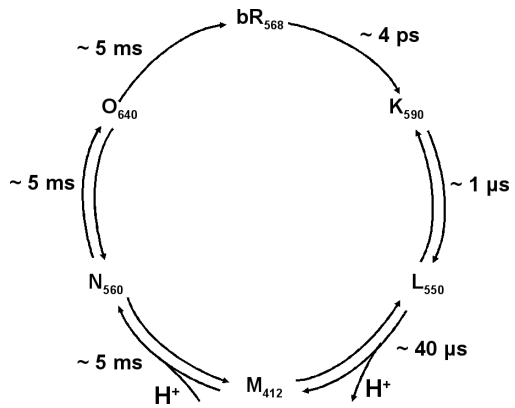


Figure 1.4: Schematic representation of the bacteriorhodopsin photocycle showing various photointermediates with their absorption maxima and their lifetimes at room temperature. The intermediates I_{460} and J_{625} are formed within first few femtoseconds after the isomerization of retinal and are not shown in the figure.

In bacteriorhodopsin, the photocycle is initiated by light absorption of the all-*trans* retinal (Figure 1.2a) and consists of several photocycle intermediate states characterized by absorption maximum of the chromophore and infrared spectra (Figure 1.4). There is no universally accepted photocycle model as the spectral features of the intermediate states are affected by several factors such as the system expression,[77] excitation conditions,[78,79] data analysis,[78] temperature, pH,[77,80] hydration[79] and existence of spectrally silent transitions.[80] Nevertheless, at physiological conditions in a widely accepted photocycle (for reviews, see refs 81, 82), the photocycle reaction sequence is given as follows (Figure 1.4):

$$bR_{568} \overset{h\nu}{\rightsquigarrow} I_{460} \rightarrow J_{625} \rightarrow K_{590} \leftrightarrow L_{550} \leftrightarrow M_{412} \leftrightarrow N_{560} \leftrightarrow O_{640} \rightarrow bR_{568}$$

Although several investigations have been focused towards understanding the isomerization process of retinal from all-*trans* to 13-*cis* in bacteriorhodopsin, the intimate details of the process are still not completely understood. Two models have been proposed to explain the process. First is a two-state model[83–85] with one electronic ground state and one excited state, and the second model is a three-state model[86–89] with one electronic ground state and two electronic excited states.

On a femtosecond (fs) time scale, the photoexcitation initiates changes in the π-electron density and bond order of the retinal chain[90] and results in formation of an excited state I_{460} species possibly *via* formation of state H which denotes the Frank-Condon excited state of bR_{568}. Vibrational spectroscopic studies[91–96] have indicated that H and I_{460} contain all-*trans* chromophore. Within approximately 500 fs, I_{460} decays through conical intersection forming a thermal intermediate J_{625} in the electronic ground state.[85,97,98] The newly formed J_{625} has been indicated to have a twisted all-*trans* retinal configuration,[93,96] although the configuration of the retinal has been controversial. Isomerization of the retinal around C_{13}=C_{14} double occurs through relaxation of the J intermediate during J→K transition. The retinal in K_{590} intermediate has twisted 13-*cis* geometry as compared to the bR state.[78,93] Using time-resolved picosecond[98] and sub-picosecond[99] spectroscopy, the J→K transition has been shown to occur within 3 picoseconds (ps). Recent vibrational spectroscopic studies reconfirmed the major events described above and proposed a new model of the isomerization process[100] in which J_{625} is denoted as an excited state.

The quantum efficiency for the isomerization had been initially measured in the range as low as 0.25[101] to as high as 0.79.[14] However revised measurements showed the yield around 0.64[102–105] and were confirmed later.[106] The energy stored in the C_{13}=C_{14} bond in K state is 11.6 $\pm$ 3.4 kcal/mol as has been shown by photocalorimetric studies,[107] while recent computer simulations indicate $\sim$7 kcal/mol energy to be stored in the retinal twist.[108]

Investigation of the K→L transition has indicated the transition to be of complex nature due to the temperature and solvent condition dependence.[109–111] The K→L transition occurs within approximately 1 microsecond (μs) time-scale[109,112] possibly *via* a spectrally silent KL intermediate[113,114] and could involve relaxation of the strained chromophore[115] such that the protein becomes ready for the first PT step. The next transition from L→M occurs on $\sim$10 μs time-scale[80] and involves the first PT from the Schiff base to a nearby Asp85 residue.[62,116–119] On a similar time-scale, under physiological conditions, a proton is released to the bulk by a group of residues known as the proton release group (PRG) whose identity has been controversial.

Since bacteriorhodopsin is a vectorial proton pump, the photocycle must involve some kind of switch mechanism that ensures the unidirectionality of the pro-

ton flow. The exact nature of this switch mechanism has also been subtle and many plausible scenarios have been proposed (for reviews see refs 120, 121). This spectrally silent transition[122] is believed to occur during early-M (or M_1)→late-M (or M_2) transition and possibly involves change in the Schiff base orientation from the extracellular to the cytoplasmic side of the membrane.[121] The switch mechanism is also believed to be associated with large conformational changes on the cytoplasmic side of the protein.[123]

The reprotonation of the Schiff base from Asp96 occurs during late-M (or M_2)→N transition.[124] Asp96 is subsequently reprotonated from the cytoplasmic bulk followed by thermal reisomerization of the retinal during N→O transition from 13-*cis* to all-*trans* configuration. Studies have indicated that these two events are coupled, based on the photocyle of Asp85Asn/Phe42Cys mutant.[125] The deprotonation of Asp85 to the extracellular PRG during final O→bR transition then completes the photocycle. Compared to the early half of the photocycle, the structural details of this last PT step are poorly understood and the transfer of proton could involve transient intermediate[126] presumably characterized by deprotonated Asp85 and protonated Asp212.

The photocycle of bacteriorhodopsin involves formation and decay of various photo-intermediates characterized by their absorption maxima. Substantial amount of discoveries have been made to investigate various aspects of the first half of the photocycle as compared to the later half. The results from various crystallographic and spectroscopic experiments are summarized in the next chapter.

Chapter 2

Discoveries in Bacteriorhodopsin I: Insight from Experiments

Summary

The bacteriorhodopsin photocycle is initiated after light-induced isomerization of the retinal and eventually one proton is transported across the membrane. The photocycle involves formation and decay of several intermediate states accompanied by PT and other structural events. In this chapter, these photocycle events are described in detail. The chapter summarizes the discoveries made from several experimental studies including structural determination methods like X-ray crystallography and spectral fingerprint techniques such as FTIR and Raman spectroscopy.

2.1 Structure Determination Methods

The pioneering work of Henderson and Unwin[24] provided first structural model of bacteriorhodopsin using electron crystallography from two dimensional crystal samples with a resolution of 7 Å with space-group P3[1]. Advances in electron crystallography for structural biology, and an inherent tendency of bacteriorhodopsin to form well ordered crystals, led to the determination of the first three-dimensional

[1] A space-group of a crystallographic group describes the inherent symmetry of the crystal structure.

structural model of bacteriorhodopsin at 3.5 Å resolution in 1990 and provided insight into the location of the amino acid side chains essential for the ion transport mechanism.[127]

With the development of cubic lipid phase crystallization method,[128] three-dimensional crystal structures of various bacteriorhodopsin photocycle intermediates were solved at improved resolutions with $P6_3$ space-group. In 3-D crystals of bacteriorhodopsin grown by lipidic cubic phase, the protein packs in layers which are stacked along the direction perpendicular to the plane of the membrane.[129] For reviews of crystallization methods and their applications to bacteriorhodopsin structure determination, see refs 35, 121, 130–132. It was soon discovered that the crystals grown the cubic lipid phase are strongly merohedrally twinned.[133,134] In spite the fact that the twinning[2] can be accounted, the effective resolutions are usually worse than that for the non-twinned crystals.[131,134,135]

The quality of the crystals also depends on the exposure of bacteriorhodopsin to detergents and crystallization in presence of lipids that are non-native to the purple membrane.[34] The diffraction quality of the sample deteriorates due to the radiation damage of the sample[130,136] and sample mosaicity.[137] The loss of electron density for carboxyl groups of acidic residues[138] due to the radiation damage is another source that affects the quality of the crystal structures. The effect of induced radiation on bacteriorhodopsin crystals indicated that the purple species (with absorbance maximum ∼570 nm) convert into orange species (with absorbance maximum ∼450-470 nm).[139]

Despite the problems mentioned above, the determination of high resolution crystal structures have provided valuable infomation regarding the structural configurations relevant for the ion transport mechanism in bacteriorhodopsin.

2.2 Photo-intermediate Trapping Protocols

Studies of bacteriorhodopsin photocycle at low temperature have revealed occurrence of two phase transitions as a function of temperature that correspond to distinct changes in the nature of allowed motions. The first phase transition occurs at 150 K, such that a change in the linear dependence of the mean square ampli-

[2] In merohedral twinning, lattices of different twin domains in a single crystal overlap in three dimensions.

tude of motion with temperature is observed.[140] FTIR spectroscopy[141] identified the second phase transition at 240 K. The studies[140,141] revealed that the deprotonated retinal, which is reprotonated from the cytoplasmic side above 240 K, becomes reprotonated from the extracellular side below 240 K. Thus, a correlation exists between the temperature at which these phase transitions occur, and the temperature at which specific intermediates are populated after bacteriorhodopsin is illuminated.

The determination of the structural features of any reaction intermediate within the crystals relies on the fact that suitable experimental conditions can be developed under which the desired intermediate can be accumulated in sufficient amounts and that the protein is functionally active in the crystalline state.[142] Various procedures for trapping structural intermediates of the light-activated systems have been employed that include: illumination of 3-D crystals at low temperatures,[143–146] room temperature reaction initiation,[147] pump-probe experiments using a broad-spectrum X-ray probe,[148,149] and to select a mutant for which the decay of the desired intermediate is delayed thus recovering a relatively high accumulation of this intermediate. Over the years several structures of bacteriorhodopsin intermediates have been proposed using above described trapping techniques.

2.3 Intermediate States and Photocycle

2.3.1 Bacteriorhodopsin Resting State (bR)

The retinal adopts an all-*trans*, 15-*anti* configuration[70] in the light-adapted bacteriorhodopsin resting state (bR; λ_{max} = 568 nm) with the Schiff base NH group oriented towards the extracellular side of the membrane.[150] Over the years several crystal structures have been proposed for bR with differences dominantly in the internal side-chains and in the loop regions.[151] The crystal structures of bR better than 2 Å resolution, in particular, structure at 1.55 Å (pdb code: 1C3W)[21] and 1.9 Å (pdb code: 1QHJ)[76] revealed structural features of several important protein groups and internal waters molecules.

The structures revealed location of a key water molecule w402 that hydrogen bonds to the Schiff base and to the anionic Asp85 and Asp212 protein groups (Figure 2.1). This water molecule w402 has been suggested to stabilize the high pK_a of the Schiff base.[152] The retinal adopts a slightly twisted all-*trans* con-

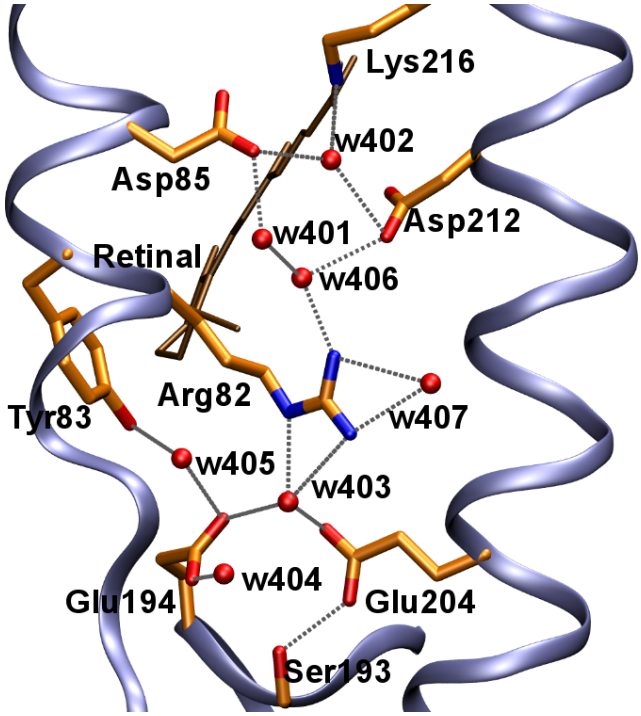


Figure 2.1: A snapshot of the extracellular region of bR indicating the hydrogen bonded network from the retinal Schiff base to Glu194/Glu204 residues *via* active site water molecules shown in red. The picture is prepared from bR structure of ref 21 (pdb code: 1C3W).

figuration[153,154] that is accommodated between tryptophan residues (Figure 1.3) that were suggested to enhance the light-induced dipole in bR and stabilize the light-induced charge distribution.[155] Two additional water molecules w401 and w406, are present in the Schiff base region and hydrogen bond the cytoplasmically oriented Arg82 with Asp85/Asp212 protein groups. Arg82 is further hydrogen bonded to w407 and to the Glu194/Glu204 protein groups via water molecule w403 (Figure 2.1). Additional hydrogen bonds are present between water molecules w404 and w405 with Glu194.

2.3.2 K State

Following the retinal isomerization, the red-shifted K state (λ_{max} = 590 nm) rises within a few picoseconds with a 13-*cis*, 14-s-*trans*,15-*anti* retinal.[78] Projection

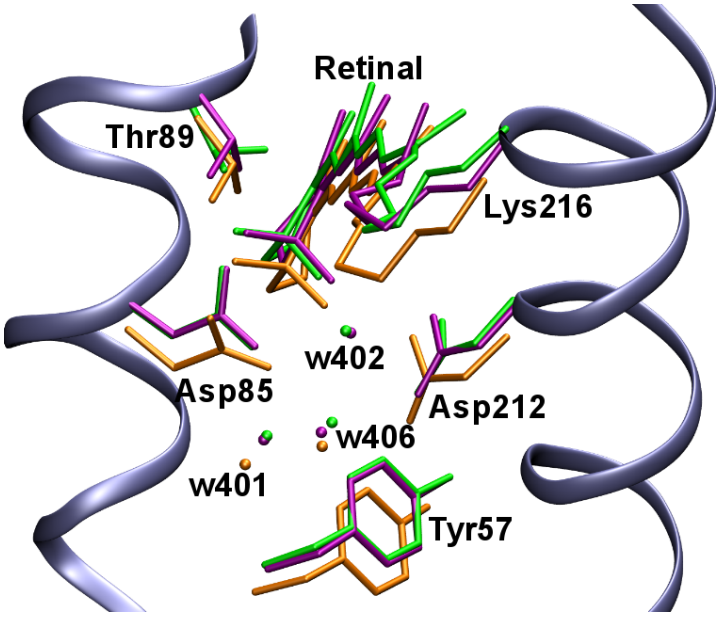


Figure 2.2: A comparison of various K state structures in the retinal Schiff base region. Color codes of the structures: 1QKO[156] in orange, 1MOK[157] in purple, and 1IXF[139] in green. Water molecules are represented as spheres and numbered according to the bR state structure of ref.[21]

map at 3.5 Å resolution obtained using cryo-electron microscopy indicated no significant structural change in the K state,[158] while the X-ray crystal structures of the K state[139,156,157] indicate rearrangements that are restricted mainly to the retinal region (Figure 2.2).

Spectroscopic studies have indicated specific and localized changes in the hydrogen bonding strengths of specific proteins groups including threonines, aspartates and tryptophans.[159–161] In particular, FTIR experiments combined with tryptophan isotope labelling observed that atleast two tryptophan residues undergo structural alterations[160] and suggested that the role of the tryptophan residues is to restrict the light-activated retinal isomerization to the C_{13}=C_{14} double bond by sterically hindering the motion of the retinylidene chromophore.[160] UV-visible absorption spectroscopy studies of site-specific mutants of bacteriorhodopsin also have supported this role of tryptophan residues.[162]

The retinal adopts a relatively planar 13-*cis*, 15-*anti* configuration with the Schiff base NH group oriented towards the cytoplasmic side in a K state struc-

ture[156] with 2.1 Å resolution (pdb code: 1QKO) derived from the crystals illuminated at 100 K. Absence of water w402 either due to its movement away from Asp85 or being disordered was suggested to allow the movement of Asp85. Further the movements of nearby protein groups including Tyr57 and main-chain Lys216 were suggested to be responsible for the initiation of the subsequent photocyle reactions.

In spite of the similar illumination protocol as of ref 156, K state structure proposed at 1.43 Å indicated significantly different structure for K, especially with respect to the active site around the retinal.[157] The highly twisted retinal with complete isomerization around C_{13}=C_{14} double bond is accompanied by partial counter-rotations of the C_{14}-C_{15} and C_{15}=N bonds such that the Schiff base is oriented towards the extracellular side. Another K state structure with 2.6 Å resolution (pdb code:1IXF) also indicated a similarly twisted retinal.[139]

In order to understand the factors that initiate the deprotonation of the Schiff base during L→M transition, it is essential to understand the geometry of the active site especially with respect to the orientation of the Schiff base in the K state i.e. an intermediate state before the L state. Except the structure of ref 156 with planar retinal, the remaining two crystal structures of refs 139, 157 indicate a twisted retinal of extracellular orientation of the Schiff base NH bond in agreement with the low temperature FTIR studies.[163] On the contrary, the picosecond time-resolved resonance Raman spectroscopy experiments suggested that relaxation in the twisted retinal is accompanied with the formation of the K state at room temperature.[93] It has been argued that these discrepancies can be attributed to the different experimental temperatures, since at low temperatures the retinal planarization may be hampered by lesser structural freedom in the protein.[93]

Information regarding the location of w402 in K state has also been contradictory. Absence of w402 in K state structure of ref 156 is consistent with FTIR studies at low temperature that indicated loss of Schiff base hydrogen bond in K.[164] On the other hand, other K state structures of refs 139, 157 indicate the presence of w402 which was also observed in FTIR studies.[159]

Another conflicting issue is the existence of Thr89:Asp85 hydrogen bond in K state. Crystal structures of refs 139, 157 indicate the presence of this bond in accordance with the FTIR studies,[164] while in the structure of ref 156, the distance of 3.8 Å between Thr89:Asp85 indicate that such a bond is unlikely.

2.3.3 KL State

Much less is known about KL state as compared to other bacteriorhodopsin photointermediate states. Nanosecond time-resolved IR studies showed that KL state builds up from the K state over ∼10 ns time.[165] The studies suggested the difference between K and KL states may be due to different chromophore twist around C_{14}-C_{15} bond, as observed by the change in the hydrogen-out-of-plane (HOOP) modes at 958 cm^{-1} and 985 cm^{-1}.[165] The 958 cm^{-1} mode was assigned[166,167] to the C_{15}HOOP and was suggested to be indicative to KL state. Other spectroscopic investigations of early bacteriorhodopsin intermediates[113,114] indicated spectral transition to formation of KL state (λ_{max} = 596 nm)[114] through decay of K state and may represent different K→KL transition than that of ref 165. On the other hand, another UV-Vis spectroscopic studies indicated K→KL transition to be spectrally silent.[168]

With the formation of KL state the Schiff base and the counterion hydrogen bond was suggested to become stronger than in previous K state.[93] Reversible transition from KL→L takes place over ∼100 ns time[165,169] and may be accompanied with relaxation in retinal twist coupled with protein structural changes.[170] During KL→L transition, structural changes of Asp96 and Asp115 carboxyl groups are coupled with the movements near the β-ionone ring of the retinal on the ∼400 ns time-scale with rearrangements near the Schiff base that exhibit in the μs range.[167] Following the observed perturbations of protein vibrational modes, it was suggested that during the K→KL and KL→L transitions, the energy stored in the distorted retinal is transferred to the protein environment.[165]

2.3.4 L State

Understanding the structural features of L state is essential for understanding the mechanism of the first proton transfer (PT) that takes during L→M transition. Consequently several experimental and theoretical studies have been dedicated to characterize features of the L state. Nevertheless precise structural details of the L state that illustrate the relative configurations of the active site protein groups including orientation of retinal Schiff base NH bond and location of active site water molecules, and consequently the mechanism of the primary PT still remain disputed.

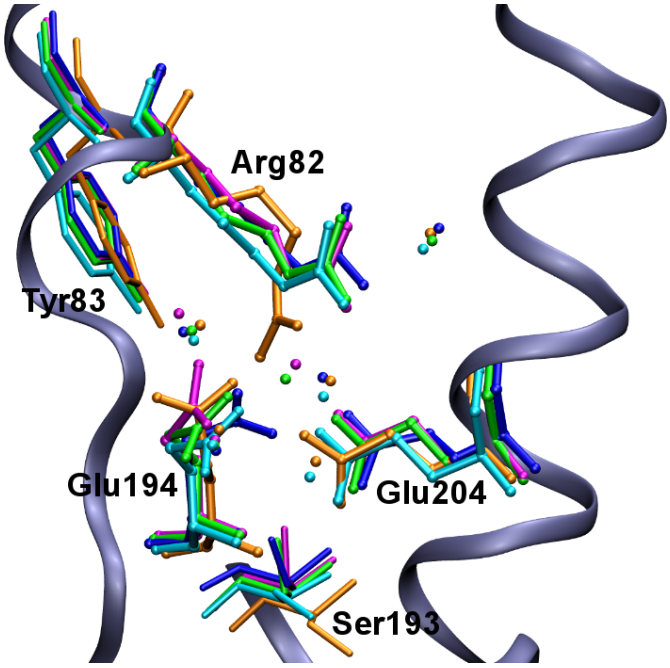


Figure 2.3: A comparison of various L state structures on the extracellular side. Color codes of the structures: 1EOP[171] in orange, 1O0A[172] in green, 1UCQ[173] in blue, 1VJM[135] in cyan, and 2NTW[174] in magenta. Water molecules are represented as spheres.

No significant protein conformational alterations take place during the formation of L state apart from observed specific structural changes around retinal active site.[123,171,172,175] Although over the years several L state crystal structures have been published, their structural information do not consistently agree with the interpretations from various spectroscopic and computer simulation studies.

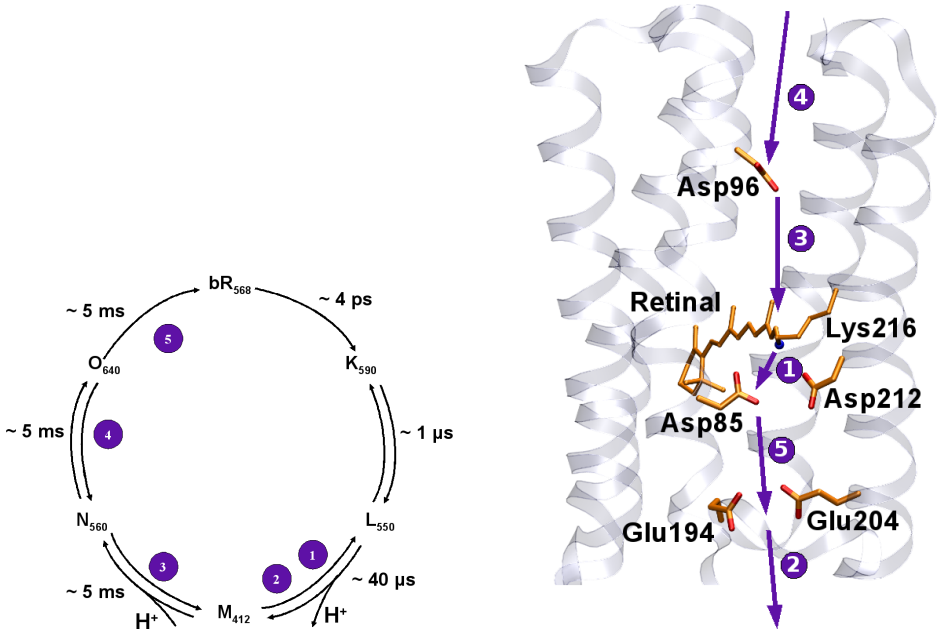
Royant et. al initially proposed[171] L state structure with 2.1 Å resolution solved at 170K (pdb code: 1EOP). Assignment of this structure to the L intermediate structure has been controversial due to contamination with the K and M intermediate states.[132,176,177] Absence of w402 molecule in this structure was argued[171] to trigger movements of Ala215 and Lys216 along with the extracellular protein groups Arg82, Tyr83, Asp85, Trp86 and Thr89. Arg82 adopts an extracellular orientation most likely due to the contribution from M intermediate and thus contradicts with other available L state structures (Figure 2.3). The structural details of the retinal region remain unchanged in the later proposed L state structure[135]

trapped at 150K (pdb code: 1VJM). The retinal is planar with cytoplasmic orientation of the Schiff base and no water molecule in the hydrogen bonding distance from the Schiff base.

Another L state structure at 1.62 Å resolution solved at 170K (pdb code: 1O0A) has twisted 13-*cis*, 15-*syn* retinal configuration with an extracellular orientation of the Schiff base.[172] Water w402 is slightly displaced and adopts a position such that it can share hydrogen bond with the Schiff base and it was proposed that w402 may act as an intermediate proton carrier.[172] The twisted 13-*cis*, 15-*syn* retinal configuration contradicts with the spectroscopic studies that suggest C_{15}=N bond to be *anti* rather than *syn*.[69,117,178] It was argued[173] that 15-*syn* retinal configuration may in fact correspond to the orange or blue absorbing species produced by X-ray absorption.[139]

Another L state structure at 2.4 Å resolution solved at 160K (pdb code: 1UCQ) indicates a 15-*anti* retinal with cytoplasmic orientation of the Schiff base.[173] Following the observation of extracellular Schiff base orientation in K state structure of ref 139, it was argued that during K→L transition, vertical movement of water molecule w602 (same as w402 of ref 21) from below to above the Schiff base is coupled with change in the Schiff base orientation from extracellular to cytoplasmic configuration.

A recently proposed L state structure at 1.53 Å resolution (pdb code: 2NTW) indicates a twisted 15-*anti* retinal with extracellular orientation the Schiff base.[174] The Schiff base is hydrogen bonded to Asp85 via water w402. However, these results are not consistent with the recent NMR[179] and FTIR studies.[180] The structure indicate the guanidinium group of Arg82 is oriented more towards the extracellular side compared to that in the bR structure of ref 21, but less than that indicated in the M state structure of ref 181. Further, the Glu194:Glu204 distance is 3.8 Å and suggest absence of hydrogen bond between Glu194:Glu204. It was argued that the lack of coupling between these two glutamates groups is due to the movement of the guanidinium group of Arg82.[174] Glu194 has a direct hydrogen bond with Tyr83 unlike in other L state structures perhaps due to the disrupted coupling between Glu194 and Glu204.



(a) (b)

Figure 2.4: A schematic representation of PT sequence (a) along the photocycle; (b) in the protein between donor and acceptor groups.

2.3.5 First PT Step: from Schiff base to Asp85

In the bacteriorhodopsin photocycle, the first PT take place from the Schiff base to a nearby Asp85 during L→M transition (Figure 2.4) over ∼10 μs time-scale[80] over ∼4 Å distance. Despite several detailed investigations, many structural and mechanistic details crucial for an understanding of the overall reaction are disputed.

The retinal configuration and orientation of the Schiff base prior to the PT has been disputed. Available crystal structures suggest diverse possibilities of the Schiff base orientation (see section 2.3.2 and section 2.3.4). The exact role of water molecules in the primary PT has also been a matter of debate. NMR experiments indicated that interactions of the Schiff base with its counterion in L state are strengthened as compared to that in the K state and suggest that a polarized water molecule can consistently explain the results and thus may act as a counterion to the Schiff base in the L state.[179] The NMR indications of a polarized water molecule led to the speculation, if bacteriorhodopsin may act as an inward

hydroxyl ion pump, rather than an outward proton pump.[182–184]

Over the years various PT scenarios have been proposed[118, 121, 132, 135, 161] based on available L and M state crystal structures and spectroscopic signatures of these intermediates: i) PT from the polarized water w402 to Asp85 followed by movement of newly generated hydroxyl ion towards the cytoplasmically oriented Schiff base to get reprotonated;[182, 185–187] ii) transfer of proton from the Schiff base to Asp85 is followed by translocation of a water molecule from the extracellular to the cytoplasmic side of the Schiff base.[173]

2.3.6 M State

The deprotonation of the Schiff base to the extracellular Asp85 is accompanied with the formation of the blue-shifted M intermediate, while the reprotonation of the Schiff base by Asp96 which is situated on the cytoplasmic side of the Schiff base, coincides with the decay of the M intermediate. Thus, the vectoriality of the proton transport is maintained by deprotonation and reprotonation of the Schiff base from the opposite sides of the membrane. Therefore, it has been suggested that more than one M state may exist that differ in the Schiff base accessesibility to protons.[186] Further, various kinetic analysis studies have also indicated that M intermediate may be composed of one or more spectrally silent substates.[111, 188, 189]

Kinetic data analysis introduced a linear photocycle model that includes an irreversible $M_1 \rightarrow M_2$ transition.[190–192] Time-resolved absorption studies indicated that the M intermediate builds up and decays with more than one time constant suggesting the possibility of more than one M states.[193] Analysis of projection maps from FTIR spectroscopy of samples illuminated at 220 K, 230 K and 240 K indicated that the M sub-states trapped at 220 K and 230 K are similar, while the M state trapped at 240 K has a different structure and FTIR spectrum.[175] Resonance Raman studies[194, 195] have also suggested similar results. In addition to these results, X-ray diffraction and FTIR experiments have suggested tertiary structural alterations take place between M sub-states.[196, 197] All these and several other results have pointed to the existence of atleast three M sub-states, although their exact number and spectral characteristics are disputed.

Several wild-type and mutant crystal structures of the M sub-states have been proposed so far. The first M state (M_1, also known as an early-M or M_L) builds up upon the Schiff base deprotonation[198] on a μs time-scale and resembles previous

L state.[186] FTIR experiments[122] have shown that the spectrally silent $M_1 \rightarrow M_2$ transition takes place on the time-scale of 350 μs and the subsequent M_2 substate builds up in $\sim$ 40 μs.[198] Enthalpy and entropic studies have also showed that $M_1 \rightarrow M_2$ transition is accompanied with major conformational changes[123,196,199–203] which may be regarded as the accessibility switch[191,204] (see section 2.3.9). Concomitant with these global protein conformational changes and local chemical changes,[205] charge redistribution has been suggested to take place around the Schiff base and on the extracellular side of the Schiff base, in presence of the cytoplasmically located anionic Asp38.[197] Eventually, M_3 (also labeled late-M or M_N) builds up in $\sim$130 μs and the equilibrium of this transition shifts towards the N state.[186,198]

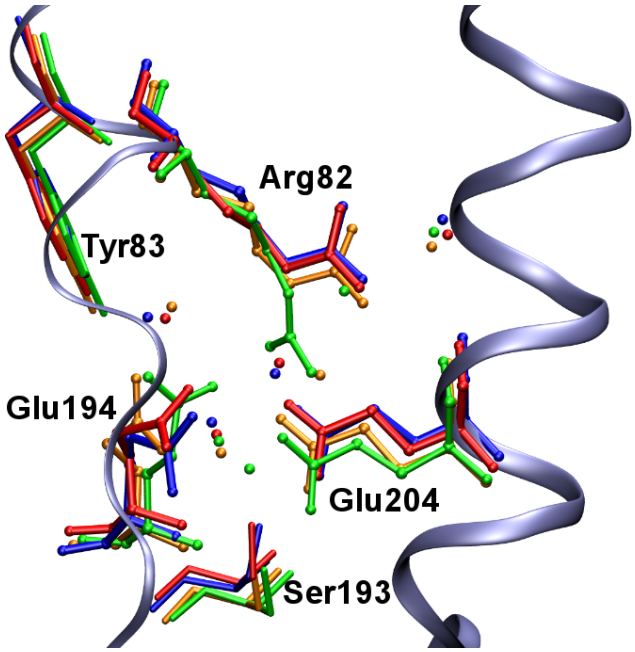


Figure 2.5: A comparison of various early-M state structures with bR (pdb code: 1C3W). Color code for pdb structures are: 1C3W[21] in red, 1M0M[206] in orange, 1KG8[207] in green and 1P8H[208] in blue.

An overlap of the so-called early-M states (pdb codes: 1M0M;[206] 1KG8[207] and 1P8H[208]) with bR structure of ref 21 is shown in Figure 2.5. All the early-M structures show a bR-like cytoplasmic Arg82 configuration except in the structure proposed by Facciotti et al (pdb code: 1KG8)[207] (see structure in green in

Figure 2.5). However, the strong hydrogen bond between Glu194/Glu204 pair observed in the bR (section 2.3.1) and in the L state structures (section 2.3.4) is absent in these structures. All the structures show an extracellular orientation of the Schiff base which is a peculiar feature of the early-M state. Nevertheless, it is not clear whether these structures represent a state before the proton release by the proton release group (PRG; see section 2.3.8) or after. This difficulty is conceivable since the first PT step (from the Schiff base to Asp85) and the second PT step (proton release by PRG) occur concomitantly under physiological conditions.[209]

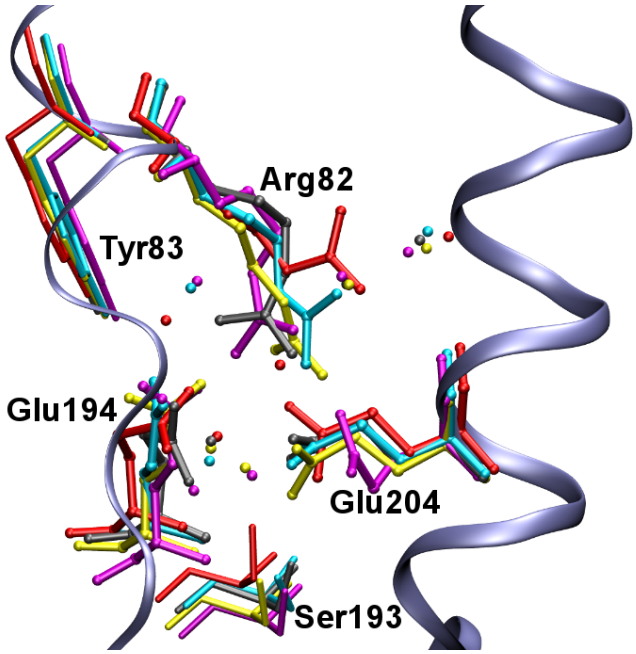


Figure 2.6: A comparison of various late-M state structures with bR (pdb code: 1C3W). Color code for pdb structures are: 1C3W[21] in red, 1C8S[210] in cyan, 1CWQ[181] in magenta, 1IW9[211] in yellow and 1F4Z[212] in gray.

An overlap of the so-called late-M states (pdb codes: 1C8S;[210] 1CWQ;[181] 1IW9[211] and 1F4Z[212]) with bR structure of ref 21 (represented in red color) is shown in Figure 2.6. In contrast to the above early-M structures, the so called late-M state structures have significant similarity in the active site (see Figure 2.6). For example, all the structures show extracellular orientation of Arg82 guanidinium group indicating a coupling between Arg82 and the extracellular Glu194/Glu204

residue pair. The Glu204Gln mutant proposed by Luecke et al[212] (represented in gray in Figure 2.6) has slightly different orientation of the Arg82 and it is due to the mutation of a nearby Glu204 residue. Assignment of M state structure of ref 181 to the late-M state has been controversial based on the indications that the magnitude of the tertiary structural changes may have been affected by crystal packing[213] and such conformational changes were not reproduced in another structural analysis of the M state.[208]

2.3.7 Movement of Arg82 Side Chain

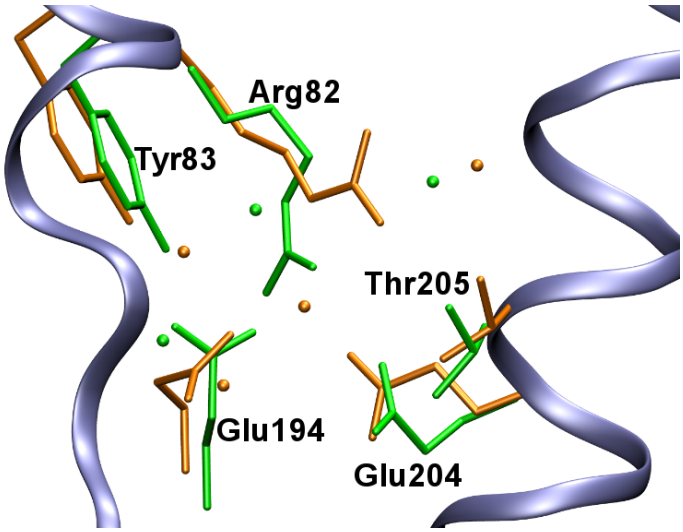


Figure 2.7: An overlap of bR structure[21](represented in orange) with a late-M state structure[181](represented in green) on the extracellular side of bacteriorhodopsin. The guanidinium group of Arg82 adopts cytoplasmic orientation in bacteriorhodopsin while following the primary PT, it adopts extracellular orientation and moves closer to the Glu194/Glu204 protein groups. Water molecules are represented as spheres.

Following the primary PT reaction from the Schiff base to Asp85, the guanidinium group of the positively charged Arg82 undergoes cytoplasmic-to-extracellular movement (Figure 2.7). The movement of Arg82 was observed when there is normal proton release to bulk (second PT step) as in the wild-type photocycle,[181] and in the photocycle of Asp96Asn mutant.[210] The movement of Arg82 was also observed when the proton release is prohibited as in the photocycle of Glu204Gln

mutant.[212] Therefore, the exact origin of the Arg82 movement is unclear.

Indeed the M state structures of wild-type bacteriorhodopsin,[181] and mutants structures of Glu204Gln,[212] and Asp96Asn[210] indicated an extracellular orientation of the guanidinium group of Arg82. These observations led to the postulation that the driving force of the Arg82 movement is the protonation of Asp85 and not the proton release to the bulk.[181,210,212] Based on the analysis of pH dependence on the photocycle of several Arg82 mutants[214] indicated that the pK_a of the PRG decreases from ∼9.5 in the bR to ∼5.8 in the M state thereby achieving the proton release and thus it was suggested that Arg82 regulates the pK_a of the PRG.

2.3.8 Second PT Step: Proton Release by PRG

In wild-type bacteriorhodopsin photocycle, concomitant to the primary PT, another proton is released to the extracellular bulk (Figure 2.4) by a group of residues known as the PRG.[209] Despite several suggestions, the identity of the PRG has been controversial. Previous studies suggested the PRG to be Glu204,[215] a complex that includes Glu194 and Glu204,[125,216,217] and a proton shared between Glu194 and Glu204.[218,219]

In yet another scenario, FTIR spectra[220] and electrostatic calculations[221] were used to postulate that the PRG consists of a protonated water cluster (i.e Zundel ion) stabilized by the Glu194, Glu204 and Arg82 residues.[222,223] The Zundel cluster would be responsible for the presence of a continuum band in the range 1800 cm^{-1} - 2000 cm^{-1}, which vanishes after the proton release during L→M transition. In contrast, other Time-Resolved FTIR experiments[224] observed a partial deprotonation of Arg82 during proton release the extracellular bulk and thus suggested the PRG to be Arg82.

Thus, to elucidate the identity of the PRG, following questions must be answered: i) if the proton is not delocalized on the water cluster as suggested by refs 222, 223, then what gives rise to the observed continuum band mentioned above, and ii) if the proton is indeed localized on either Glu194 or Glu204, then why no change in the characterstic ν_{COOH} bands are observed in the FTIR studies?

It is also unclear if the proton release is coupled to the Schiff base deprotonation during the primary PT reaction. Titration analysis of Asp85 in the non-illuminated state has indicated that the protonation of Asp85 and the PRG are coupled.[225] Site-directed mutagenesis and subsequent flash photolysis and pH indicator experiments

have indicated that a coupling between Asp85, Asp212 and Arg82 is essential for the proton release.[226]

2.3.9 The Accessibility Switch

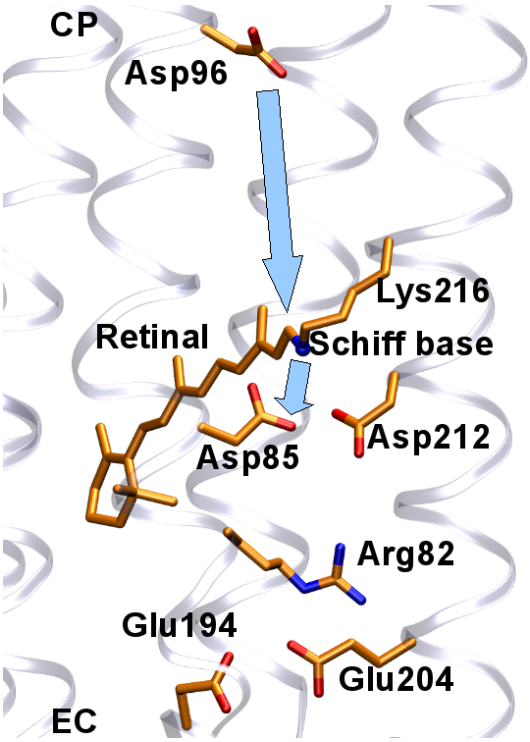


Figure 2.8: Structure of bacteriorhodopsin in the bR taken from ref 21. The cytoplasmic and extracellular side are denoted as CP and EC respectively. The Schiff base deprotonates to Asp85 situated on the EC side of the Schiff base during L→M transition, while receives a proton from Asp96 situated on the CP side during M_3→N transition.

To actively pump ions across the membrane, an ion transporter protein requires to switch access from one side of the membrane to other, but not to both at the same time.[82,193,227–229] In bacteriorhodopsin, the Schiff base deprotonates to Asp85, located on the extracellular side of the retinal while gets reprotonated from Asp96 which is located on the cytoplasmic side of the retinal during M_3→N transition (Figure 2.8). This change in the Schiff base orientation such that it receives protons from different sides of the membrane is generally known as reprotonation (or

protonation) switch.[81,127,178,191,204,227,229–231]

Changes in the infrared bands of the slowly decaying M state in the temperature range 240-260 K indicated that the Schiff base becomes reprotonated from Asp85, located on the extracellular side of retinal and emphasized the importance of protein conformational changes for the vectoriality of the proton transport.[141] The investigation of the protein conformation changes during the photocycle of Asp96Gly/Phe171Cys/Phe219Leu triple mutant,[123,213] led to the suggestion that the intermediate structures of bR, K, L, and M_1 can be represented by one protein conformation such that the Schiff base is preferentially accessible to the extracellular side of the retinal. On the other hand, the intermediate structures of the later half of the photocycle (M_2, N and O) can be represented by another protein conformation where the Schiff base accessibility has switched to the cytoplasmic side of the retinal. However, the fact that this triple mutant exhibits normal proton pumping activity[232] makes the proposition of the switch mechanism based only on global conformations questionable.

Various other possibilities for the switch mechanism have also been proposed over the years: (i) following the proton release by the PRG to the extracellular bulk, increase in the pK_a of Asp85 was proposed such that its probability as a proton donor to the Schiff base decreases as compared to Asp96 located on the cytoplasmic side;[225,233] (ii) a movement of the Schiff base away from Asp85 was proposed following the deprotonation of the Schiff base with a change in the curvature of the retinal polyene chain;[213] (iii) an electrostatic steering mechanism was suggested that involves reorientation of the Schiff base NH group from the extracellular to the cytoplasmic side after breaking of its electrostatic interaction with Asp85,[183] consistent with the observed X-ray structures of M_1 and M_2 states;[206] (iv) a movement of the segment of helix C was proposed such that prior to the PT i.e. in L state, Asp85 moves towards the Schiff base and after the PT i.e in M (or early-M), moves away from the Schiff base;[171] (v) a model was proposed that suggested the function of bacteriorhodopsin as an inward hydroxyl ion pump rather than conventionally believed proton pump.[182] This model also suggested that following the PT step from the Schiff base to Asp85 involving dissociation of the water molecule w402, the newly created hydroxyl ion would move from the extracellular to the cytoplasmic side where it would function as a proton acceptor;[182] (vi) the vectoriality of the PT was suggested to be due to the spatial separation between the

Schiff base and Asp85;[132] (vii) based on the FTIR experiments, a hydration switch model was proposed which suggested that increase in the pK_a of protonated Asp85 is accompanied with the movement of an active site water molecule.[234]

Thus, although the necessity of the protonation switch is clear, the precise nature of this switch has remained controversial.

2.3.10 Third PT Step: Reprotonation of the Schiff Base

Following the accessibility switch event(s), the Schiff base receives a proton from the cytoplasmically oriented Asp96 (Figure 2.4) during $M_3 \rightarrow N$ transition (Figure 2.8). FTIR studies indicated that appearance of the negative band at 1742 cm^{-1} indicative of the deprotonation of Asp96, coincides with the reprotonation of the Schiff base.[118,235–239] Further studies[235,236,240] in which Asp96 was mutated also supported the role of Asp96 as an internal proton donor to the Schiff base. The transfer of proton from Asp96 to the Schiff base takes place over a distance $\sim$11 Å and requires decrease in Asp96 pK_a from $\sim$11 in bR state[241] to $\sim$8 in N state.[242] The retinal isomerization during bR$\rightarrow$K induces an initial strain that was suggested cause protein conformational changes coupled to the deprotonation of Asp96.[243] These studies also indicated that when the donor Asp96 is inhibited as in case of Asp212Asn/Asp96Asn mutant, the Schiff base is reprotonated directly from the cytoplasmic bulk but with much slower rate.[243]

In contrast to previous suggestions,[238,244] minor conformational changes that take place during the rise of N state[123,205] may be responsible for generation of a water file. The formation of such a water file through which the proton from Asp96 is carried to the Schiff base by a Grotthus mechanism has been suggested in several studies.[245–253] The formation of a water file is perhaps due to the strong influence of osmotic-,[246] and hydrostatic pressure[248] on the $M_3 \rightarrow N$ transition relative to any other reactions in the photocycle.

2.3.11 N State

The N state is the only state in which Asp96 is anionic. The retinal adopts 13-*cis*, 14-s-*trans*, 15-*anti* configuration[178] in N state (λ_{max} = 520 nm). The strength of hydrogen bond between the retinal and protein remains unaltered between L and N intermediates states despite the fact that retinal twist is reduced in the N state than

in the L state.[254]

Large conformational changes were observed in the N state structure of Phe219Leu mutant[244] solved using electron microscopy at resolution 3.5 Å. The involvement of the E-F loop in large conformation changes has also been suggested to take place during $M_3 \rightarrow$N transition.[255] Although these results support the suggestion of large scale conformational changes during $M_3 \rightarrow$N transition,[213] the N state structure derived from Asp96Gly/Phe171Cys/Phe219Leu triple mutant suggested that minor structural changes are associated with the formation of N state.[123,205] These suggestions were based on the observed similarity between the conformational changes associated with the N state and those observed in the late M state.[123]

Although no wild-type N crystal structure has been proposed so far, the structure of a late-N state that corresponds to a state after reprotonation of both Asp96 and the Schiff base has been reported with 1.62 Å resolution.[208] The retinal in this late-N state is in relaxed 13-*cis*, 15-*anti* configuration with cytoplasmically oriented Schiff base NH group. The structure shows a hydrogen bonded chain of water molecules from Asp96 to the Schiff base and thus support a possibility that the PT from Asp96 to the Schiff base may involve intermediate water molecules.

2.3.12 Fourth PT Step: from Cytoplasm to Asp96

The fourth PT involves reprotonation of Asp96 from the cytoplasmic bulk (Figure 2.4) during N→late-N transition.[194,236,239,256] During the PT step, the pK_a of Asp96 is ∼7, presumably due to its increased connectivity to the cytoplasmic surface.[219,243,257,258] The PT pathways for the reprotonation of the Schiff base during $M_3 \rightarrow$N and for the reprotonation of Asp96 during N→late-N have been suggested to be different.[259] Mutations of several protein groups in the vicinity of Asp96 such as Thr46Val, Leu100Cys, Phe171Cys, Leu223Cys cause acceleration of the Schiff base reprotonation and deceleration of the Asp96 reprotonation.[260,261]

At the cytoplasmic surface of bacteriorhodopsin, six acidic residues, Asp36, Asp38, Asp102, Asp104, Glu161 and Glu166 are present which are crucial for the proton update and have been suggested to act as a proton funnel.[262–264] Mutation of Asp38 by Arg dramatically slows down the decay of M intermediate,[264] indicating its importance to the reprotonation of Asp96 presumably as an intermediate proton carrier.[265] In contrast, other mutation studies in which Asp38 was mutated either by Asn,[266] or by Cys[267] indicate minor alterations in the photocycle.

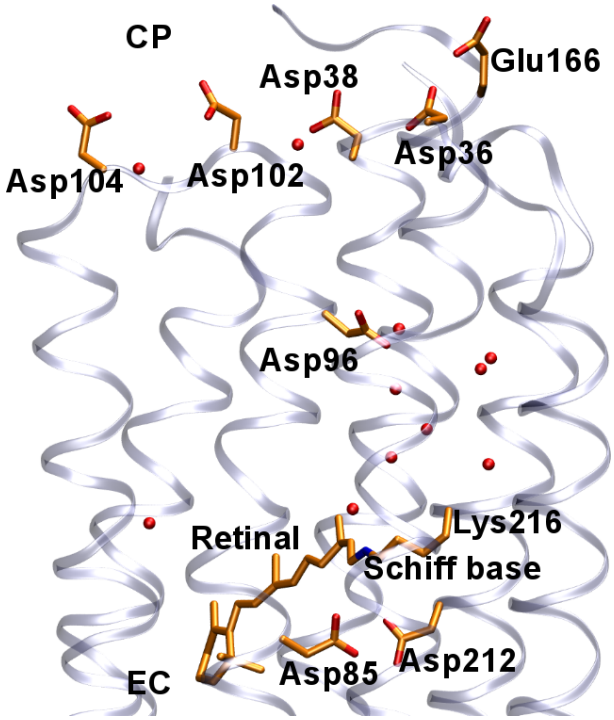


Figure 2.9: Location of acidic residues at the cytoplasmic side of bacteriorhodopsin crucial for the proton uptake by Asp96 according to the late-N state structure of ref 208. CP and EC indicate the cytoplasmic and extracellular side. Asp96 receives proton from the cytoplasmic bulk during N→late-N transition.

Although, the details of the proton uptake step by Asp96 are unclear,[258] the reprotonation of Asp96 may occur *via* a different channel than that for the retinal reprotonation and may require rearrangements of hydrophobic groups on the cytoplasmic side[258] and of hydrogen bonded groups.[268] These results support the conclusion that on the cytoplasmic side of the protein, PT from Asp38 to Asp96 would take place by transient opening and subsequent solvation of a channel connecting these two groups.[265]

The protonation states of Asp85 and Asp96 are important for the retinal geometry in N.[258,259] Behavior of Asp85Asn/Phe42Cys double mutant suggested that the reprotonation of Asp96 and back-isomerization of retinal are coupled.[125] The *thermal isomerization* of retinal back to initial all-*trans* configuration takes place during late-N→O transition.[88] The protein groups such as Leu93 are essential for

restricting the flexibility of retinal which is crucial for the retinal isomerization,[269] as well as for the coupling between the retinal geometry and changes in the protein conformation during decay of N intermediate.[270,271]

2.3.13 O State

Following the thermal reisomerization of the retinal, O state (λ_{max} = 640 nm) builds up and is characterized by all-*trans* retinal.[272,273] Resonance Raman studies have indicated that the newly formed retinal is twisted, with a magnitude of twist similar to that of the K state structure.[272] No wild-type crystal structure exists for the O state and therefore understanding the structural features of the O state has been minimal.

Two putative O-like structures have been proposed: for the chloride pumping Asp85Ser mutant[274] (pdb code:1JV7), and for the acid-blue bacteriorhodopsin[275] (pdb code:1X0I). The assignment of these structural models to an O-like state was based on the main characteristics of the O state - neutral Asp85, optical absorption maximum red-shifted relative to the wild-type bR, and the all-*trans* retinal. A recent EPR study[276] suggested that the structural characteristics of the O state protein in the purple membrane environment are better reproduced by the acid-blue bacteriorhodopsin[275] than by Asp85Ser.[274] Nevertheless, many structural details in the O state such as orientation of Arg82, number and orientation of water molecules remain unclear.

2.3.14 Fifth PT Step: from Asp85 to PRG

Recovery of the initial bR occurs through the PT from Asp85 to the extracellular PRG (Figure 2.4) over a distance of ∼12 Å over millisecond time-scale.[194] Little is known about the mechanism of this last PT step that presumably requires involvement of both extracellular Arg82 residue,[226] and an interaction between the retinal geometry with surrounding protein groups.[277] Understanding the mechanism of this final PT step requires understanding of several key issues such as: (i) whether the PT occur in a single step, or it involves intermediate(s), and (ii) what are the structural elements critical for PT from Asp85 to the PRG.

Chapter 3

Discoveries in Bacteriorhodopsin II: Insight from Simulations

Summary

In the last chapter, results from the experimental studies of bacteriorhodopsin were described. In this chapter, results from the computer simulation studies are summarized. The findings from the simulation studies are organized according to the photocycle events. Section 3.2 outlines the computational challenges and unresolved issues in the bacteriorhodopsin photocycle.

3.1 From Structure to Mechanism

Early computer simulations investigated the retinal geometry and its response to the protein environment for early bacteriorhodopsin intermediates.[278–282] Systematic investigation of the nature of the retinal geometry in bR indicated that the twist in the all-*trans* retinal is mainly around C_{13}=C_{14} and C_{15}=N_{16} double bonds.[153] The twist was suggested to elevate the pK_a of the Schiff base in bR prior to the PT.[153]

The light-activated retinal isomerization around C_{13}=C_{14} double bond alongwith the isomerization around C_{14}-C_{15} single bond was indicated to be energetically feasible and depends crucially on the protonation state of the Schiff base.[283,284] The isomerization process involves a partial rotation around several bonds of the polyene chain[285] with a time-scale of $\sim$200 fs while another quantum dynamics studies indicated a time-scale of $\sim$300 fs for the isomerization.[286]

The formation of the excited state J_{625} following the retinal isomerization was suggested to be due to a polarization of the protein environment induced by changes in the excited state charge distribution of the retinal.[287] It was further suggested that the rise of K_{590} is due to the vibrational relaxation of retinal.[287]

Starting from the electron microscopy structure of bacteriorhodopsin,[127] MD simulations[249,288] investigated the protein and retinal interaction with respect to the retinal isomerization process and suggested all-*trans*→13-*cis*, 14-s-*cis* as the retinal isomerization process. These studies also revealed the importance of Asp85-Asp212 counterions for the retinal stereochemistry. Subsequent MD simulations[250] indicated that the counterion to the Schiff base is composed of Asp85 and Asp212 connected through a hydrogen bonded network involving six water molecules, Tyr57, Arg82, Thr89, and Tyr185.

MD simulations investigated the stability, and interactions of water molecules with the protonated Schiff base in bR, and indicated that the water molecules can interact with the Schiff base in two distinct sites: one near the NH group of the Schiff base on the extracellular side of the retinal, and the other near a retinal CH group directed towards the cytoplasmic side.[289]

Following the retinal isomerization during bR→K transition, active proton transport along the photocycle would involve the conversion of the K intermediate to the subsequent L intermediate. Indeed minimum energy path calculations based on QM/MM simulations[108] indicated that the rate-limiting barrier for the forward K→L transition is ∼3 kcal/mol. In contrast, the rate-limiting barrier for the back-isomerization of the retinal that would involve K→bR transition is ∼11 kcal/mol and arises due to the energy stored in retinal.[108] Thus, the studies revealed how the protein favors the forward K→L transition for active proton pumping and prohibits the back-isomerization.[108]

Prior to the deprotonation of the Schiff base, its pK_a is lowered.[278,284,290–293] The importance of water molecules in stabilization of the protonated Schiff base and counterion configurations in the retinal active site has been suggested earlier.[289,294] Quantum chemical calculations also revealed that the hydrogen bonded network around the retinal active site controls the pK_a of the donor Schiff base and acceptor Asp85 groups prior to the PT.[295] The coupling between the extracellular PRG and the Schiff base has been suggested based on the multiconformation continuum electrostatics, and equilibrium proton distribution studies.[296] However,

these findings contradict with another theoretical study that ruled out any coupling between the first and the second PT steps.[209]

Water molecules in the active site around the Schiff base and Asp85 have substantial impact on energetics of the primary PT.[294,297,298] Recent free energy simulations[299] have indicated that prior to the primary PT reaction, the presence of a water molecule on the cytoplasmic side of the Schiff base is favorable and the water is likely to act as an intermediate proton carrier in the PT reaction from the Schiff base to Asp85.

Simulation studies using QM/MM approach[119,300] investigated various proton transfer pathways motivated from different active site geometries and indicated that three different PT pathways may coexist and the enthalpic rate-limiting energy barriers of these pathways are $\sim$11.5-13.6 kcal/mol in agreement with experiments.[80] The studies[119,300] indicated that the retinal with 13-*cis*, 15-*syn* geometry does not actively pump protons and following PT scenarios were proposed:[119,300] a) direct PT from the Schiff base to Asp85; b) Schiff base first deprotonates to Thr89 which concomitantly deprotonates to Asp85 and, c) an initial PT from Schiff base to Asp212, followed by a concerted PT from Asp212 to Asp85 via an intermediate water molecule w402.

Recent free energy simulations investigations[301] of the primary PT path indicate that the proton is also likely to be transferred directly from the retinal Schiff base to Asp85 and the overall reaction is driven by light-activated charge separation[285,302] between donor Schiff base and acceptor Asp85. Such a direct PT between the donor and acceptor groups is in agreement with earlier studies.[300] To ensure the newly protonated Asp85 remains neutral i.e. to inhibit the back PT from Asp85 to the Schiff base, likely crucial factors are: changes in the Thr89:Asp85 distance, and movement of water molecules in the active site.[298]

During the L$\rightarrow$M transition, concomitant to the primary PT reaction, the PRG releases a proton to the extracellular bulk. This identity of the PRG has also been extensively investigated by several theoretical and experimental studies. Continuum electrostatics calculations postulated[221] the PRG to consists of a protonated water cluster (i.e Zundel ion) stabilized by the Glu194, Glu204 and Arg82 and could give rise to a continuum band in the same region as observed in experiments.[220] Car-Parrinello MD studies[303,304] showed that such a protonated Zundel cluster can give rise to an IR continuum band in the 1800 cm^{-1} -2000 cm^{-1} re-

gion. However, the Zundel cluster in these simulations was stabilized by imposing artificial external constraints on solvating water molecules.[303] Further, the stability of such a Zundel cluster in the protein remains unclear.

MD simulations of the M_{412} state suggested that prior to the deprotonation of Asp96 to the Schiff base, the Schiff base nitrogen changes its orientation EC side to CP side towards Asp96, and two additional water molecules move into the cytoplasmic side of the protein such that Asp96 is connected to the Schiff base through a chain of hydrogen bonded water molecules.[305] The reprotonation of the Schiff base from Asp96 during $M_3 \rightarrow N$ transition requires a PT over $\sim$11 Å distance presumably *via* a transient formation of a chain of water molecules. Free energy simulations investigated the thermodynamic stability of internal water molecules in bacteriorhodopsin and suggested that the transfer of four water molecules from the cytoplasmic bulk into the bacteriorhodopsin cavity is thermodynamically allowed, such that the donor Asp96 and the acceptor Schiff base are connected through a chain of atleast four hydrogen bonded water molecules.[251] The formation of such a water file through which the proton from Asp96 is carried to the Schiff base by a Grotthus mechanism has also been suggested in several other theoretical studies.[249–253]

The thermal reisomerization of the retinal from 13-*cis*→all-*trans* configuration takes place during N→O transition. Classical free energy simulations indicated that the free energy barrier associated with the reisomerization process is $\sim$14 kcal/mol.[306]

Recent theoretical modelling of O state indicate that Arg82 adopts an extracellular orientation in O state and a chain of hydrogen bonded water molecules connects the donor Asp85 to the acceptor PRG.[307]

3.2 Computational Challenges and Unresolved Issues

Bacteriorhodopsin is perhaps the most widely investigated ion pump both by experimental and theoretical methods. Simultaneous to the experimental discoveries, development of advanced computational methods and powerful computer resources have elucidated the complex light-activated photocycle mechanism to an unprecedented level of detail. Nevertheless, many important details in the proton pump mechanism of bacteriorhodopsin still remain unanswered.

Many wild-type and mutant structures have been proposed from bR to M intermediate states. Although these structures differ in few structurally intimate issues, the structural models reveal locations of many protein groups and water molecules. Spectral signatures of characteristic protein groups in wild-type and mutant intermediates have been investigated detail by employing a variety of spectroscopic techniques including resonance Raman, FTIR, and NMR. This may be a reason why most of the computer simulation studies have been dedicated to understanding the early intermediate states bR, K, and L and the photocycle reactions amongst them. It is essential to note that this does not necessarily mean that the processes of the early half of the photocycle are easy to simulate.

Events that take place during the later half of the photocycle are relatively less investigated by experimental techniques especially by structural determination methods. While many wild-type crystal structures exist for early intermediates, wild-type structures for late-M, N and O intermediates have not been resolved. Hence, the only clues to the structural features of these states come from mutant structures or crystals grown under acidic conditions. This is mostly due to the difficulties in accumulating sufficient amount of these intermediate states necessary for refinement of the structure. The problem is more complicated by the possible presence of one or more substates which are perhaps through spectrally silent transitions. Consequently structural characteristics of N and O intermediates are poorly understood. Complexity of these so called *late* events is also elevated by the fact that M$\rightarrow$N, N$\rightarrow$O, and O$\rightarrow$bR transitions take place over ms time-scale and involve PT reactions over $\sim$10-12 Å distance.

Although simulation techniques can robustly investigate complex biochemical processes that involve protein conformational changes, entropic and solvation effects, their applicability is still limited by large sampling time. Perhaps this is a reason why very few studies were focused to investigate these *late* events.

To summarize, the *late* events are challenging to investigate due to (i) absence of relevant crystal structures, and (ii) significant protein conformational changes that dominate the entropic and solvent effects. Perhaps this is reason why a very few simulations studies have been done to study these events. In the present work, multi-length-scale simulations were carried out to investigate these *late* photocycle events which take place on the EC side of the protein.

Chapter 4

Theoretical Methods I: Quantum and classical Approaches

Summary

This chapter introduces the theoretical methods that have been developed over the years and which are relevant to the current work. Starting from the historical developments that led the foundations of quantum mechanics, the foundations of molecular orbital theory are summarized including *ab initio* and semi-empirical approaches. After introducing the essential developments of density functional theory in section 4.3, an approximate DFT method, the self-consistent density functional tight-binding (SCC-DFTB) is described. In the following and final section, the empirical force field approaches are summarized.

4.1 Prelude to Quantum Mechanics

The development of the laws of motions by Sir Isaac Newton in the late seventeenth century provided rigorous mathematical framework for explaining and understanding the dynamical properties of macroscopic objects. His famous three laws of motion in addition to the concept of gravity successfully provided accurate description of motion of both celestial and terrestrial objects. During the early nineteenth century, Thomas Young successfully demonstrated the wave nature of light and around 1860, James Maxwell unified the laws of electricity and magnetism into four equations now known as Maxwell's equations. In addition to

these developments, numerous other discoveries and inventions provided more and more insight into phenomena observed in nature. Thus by the end of the nineteenth century physicists had started to believe that the theoretical foundation of the laws of physics (so called classical physics) is rigorous and sufficient to explain all or nearly all phenomena observed in nature.

Despite its success, the laws of classical physics failed to explain many important experimental results. One such example is the property of black-body radiation. The laws of statistical mechanics and electromagnetic theory completely failed to explain the black-body radiation in the high frequency region. In 1900, Max Planck provided successful explanation of the complete black-body radiation using the revolutionary idea of quantized nature of light. Shortly afterwards in 1905, Albert Einstein applied the concept of quantization of light assuming the particle-like nature of light to explain the photoelectric effect.

During the following years, a revolutionary theory known as quantum mechanics was developed by many physicists including Wolfgang Pauli, Werner Heisenberg, Max Born, Erwin Schrödinger and Paul Dirac. Although the theory was based on many abstract mathematical concepts, it provided accurate and quantitative explanation of properties of microscopic particles and turned out to be a generalization of Newton's classical mechanics.

4.2 The Basics of Molecular Orbital Theory

In classical physics, Newton's equations describe time evolution of any macroscopic body under influence of a constantly acting force. In the same spirit in 1926, Erwin Schrödinger invented a time-dependent equation for microscopic particles given by,

$$i\hbar\frac{\partial\Psi(\mathbf{r},t)}{\partial t} = -\frac{\hbar^2}{2m}\nabla^2\Psi(\mathbf{r},t) + V(\mathbf{r},t)\Psi(\mathbf{r},t) \tag{4.1}$$

where ∇^2 is the Laplacian operator. The state function or wave function $\Psi(r,t)$ of a system contains every possible information about the system and depends on 3-dimensional Cartesian coordinates $\mathbf{r}(x,y,z)$ and time t. $V(\mathbf{r},t)$ is the potential energy of the system. Max Born shortly afterwards interpreted $\mid \Psi(\mathbf{r},t) \mid^2$ as the probability density for finding the particle at various places in the cartesian space. Thus, $\mid \Psi(\mathbf{r},t) \mid^2 d\mathbf{r}$ gives the probability of finding the particle at time t in the

volume element $d\mathbf{r}$ and H is the Hamiltonian of the system given by,

$$H = -\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{r},t) \tag{4.2}$$

where first and second terms on the right hand side (r.h.s.) of the above equation are the kinetic energy and the potential energy functions respectively and $\hbar$ is the Planck's constant divided by 2π.

For cases where the Hamiltonian H of the system is independent of time t, the time-dependent Schrödinger equation (Eq. 4.1) reduces to the time-independent equation given by,

$$-\frac{\hbar^2}{2m}\nabla^2\Phi(\mathbf{r}) + V(\mathbf{r})\Phi(\mathbf{r}) = E\Phi(\mathbf{r}) \tag{4.3}$$

or more commonly,

$$H\Phi(\mathbf{r}) = E\Phi(\mathbf{r}) \tag{4.4}$$

where E is the energy of the system and the Hamiltonian H is given by,

$$H = -\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{r}) \tag{4.5}$$

The Hamiltonian for a molecule with many atoms (lets say N electrons and M nuclei) is given by,

$$H = -\sum_i^N \frac{\hbar^2}{2m_\mathrm{e}}{\nabla_i}^2 - \sum_k^M \frac{\hbar^2}{2m_k}{\nabla_k}^2 - \sum_i^N\sum_k^M \frac{e^2 Z_k}{r_{ik}} + \sum_{i<j}^N \frac{e^2}{r_{ij}} + \sum_{k<l}^M \frac{e^2 Z_k Z_l}{r_{kl}} \tag{4.6}$$

where, i and j range over the electrons of the system, k and l range over the nuclei of the system, m_e is the mass of the electron, m_k is the mass of the nucleus k, e is charge of the electron, Z is an atomic number, and r_{ab} is the distance between particles a and b.

The first two terms on the right hand side of Eq. (4.6) are the kinetic energy terms of electrons and nuclei respectively. Third term is the attractive interaction term between negatively charged electrons and positively charged nuclei. The next two terms are electron-electron and nuclei-nuclei repulsion terms respectively.

For one electron systems such as for hydrogen, He^+, Li^{2+} and so on, the Schrödinger equation can be solved exactly to get the energies and wavefunctions. However for systems with more than one electron (N-electron systems), some approximations must be applied to solve the Schrödinger equation due to the complex nature of the interelectronic repulsion terms.

4.2.1 Born-Oppenheimer Approximation

In reality, solving the Schrödinger equation for molecules can be extremely complicated due to the inter-dependence of electronic and nuclear motions indicating that no particle can move independent of all others. This difficulty can be circumvented by introducing Born-Oppenheimer approximation. Under normal physical conditions, the electrons being about 1800 times lighter than the nuclei, move at much higher speeds compared to the nuclear motion. Hence, for practical purposes, the nuclear position can be assumed to be fixed during the electronic motion. This significantly simplifies the situation since, the nuclear kinetic energy term can be taken to be independent of electrons and the electron-nucleus correlation term can be eliminated. The repulsive nuclear-nuclear term (last term on r.h.s. of Eq. 4.6) thus becomes a constant term for a particular configuration of nuclear positions. The Schrödinger equation for a multi-electron system then reduces to,

$$H_{\mathrm{el}}\Phi_{\mathrm{el}} = E_{\mathrm{el}}\Phi_{\mathrm{el}} \tag{4.7}$$

where, the subscript "el" denotes the electronic Schrödinger equation and emphasizes the Born-Oppenheimer approximation. The electronic Hamiltonian H_{el} parametrically depends on the nuclear coordinates and it is given by (in atomic units),

$$H_{\mathrm{el}} = -\sum_{i}^{N}\frac{1}{2}{\nabla_i}^2 - \sum_{i}^{N}\sum_{k}^{M}\frac{Z_k}{r_{ik}} + \sum_{i}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}} \tag{4.8}$$

The electronic wave function Φ_{el} explicitly depends on the spatial coordinates of the electrons. Since for a complete description of an electron, specification of spin function in addition to the spatial coordinates are necessary. Both these variable can be collectively represented by $\mathbf{x}$,

$$\mathbf{x} = \{\mathbf{r}, \omega\} \tag{4.9}$$

where the spin coordinate ω corresponds to either an up spin function, $\alpha(\omega)$ or a down spin function, $\beta(\omega)$. Pauli's exclusion principle for an N-electron wave function is given by,

$$\Phi(\mathbf{x}_1,\ldots,\mathbf{x}_i,\ldots,\mathbf{x}_j,\ldots,\mathbf{x}_N) = -\Phi(\mathbf{x}_1,\ldots,\mathbf{x}_j,\ldots,\mathbf{x}_i,\ldots,\mathbf{x}_N) \tag{4.10}$$

Thus, the exact wave function must satisfy the Schrödinger equation as well as Pauli's exclusion principle.

An orbital is defined as a wave function for a single particle, say an electron. However, since a complete description of an electron requires specification of spatial as well as spin coordinates (Eq. 4.9), the wave function for an electron is represented as a spin orbital, $\chi_i(\mathbf{x}_i)$ which is a product of a spatial orbital $\psi_i(\mathbf{r})$ and one of the spin orbitals $\alpha(\omega)$ or $\beta(\omega)$.

4.2.2 The Hartree-Fock Self-Consistent-Field (SCF) Method

For a system of N non-interacting electrons, the Hamiltonian of the system can be written as,

$$H = \sum_{i=1}^{N} h_i \tag{4.11}$$

and the one-electron Hamiltonian h_i consists of kinetic and potential energy of electron i (neglecting electon-electron repulsion). The eigenfunctions of h_i can be a set of spin orbitals $\{\chi_j(\mathbf{x})_i\}$ with an spin orbital energy, ε_j and satisfy,

$$h_i\chi_j(\mathbf{x}_i) = \varepsilon_j\chi_j(\mathbf{x}_i) \tag{4.12}$$

Since H is a sum of one-electron Hamiltonians, an eigenfunction of H satisfies,

$$H\Phi^{\mathrm{HP}} = E\Phi^{\mathrm{HP}} \tag{4.13}$$

where eigenvalue E is sum of the spin orbital energies such that,

$$E = \varepsilon_i + \varepsilon_j + \cdots + \varepsilon_N \tag{4.14}$$

and the many-electron wave function Φ^{HP} is termed as a Hartree product given by,

$$\Phi^{\mathrm{HP}} = \chi_i(\mathbf{x}_1)\chi_j(\mathbf{x}_2)\ldots\chi_k(\mathbf{x}_N) \tag{4.15}$$

Note that each of these spin molecular orbitals (MO) depend on the coordinates of one electron which consists of spatial $\mathbf{r}$ and a spin ω variable (Eq. 4.9). Since the above defined Hartree product is symmetric with respect to an interchange of coordinates between two electrons, the requirement of the antisymmetry principle can be satisfied by defining a wave function as an appropriate linear combination two Hartree products. For two electrons, the antisymmetric wave function in a determinant form can be written as,

$$\Phi^{\mathrm{SD}} = \frac{1}{\sqrt{2}}\begin{vmatrix} \psi_i(\mathbf{x}_1)\alpha(\mathbf{x}_1) & \psi_j(1)\beta(\mathbf{x}_1) \\ \psi_i(\mathbf{x}_2)\beta(\mathbf{x}_2) & \psi_j(\mathbf{x}_2)\alpha(\mathbf{x}_2) \end{vmatrix} \tag{4.16}$$

where "SD" stands for a Slater determinant. For an N-electron system, the Slater determinant can be written as,

$$\Phi^{\mathrm{SD}} = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_i(\mathbf{x}_1) & \chi_j(\mathbf{x}_1) & \cdots & \chi_k(\mathbf{x}_1) \\ \chi_i(\mathbf{x}_2) & \chi_j(\mathbf{x}_2) & \cdots & \chi_k(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_i(\mathbf{x}_N) & \chi_j(\mathbf{x}_N) & \cdots & \chi_k(\mathbf{x}_N) \end{vmatrix} \tag{4.17}$$

The N-electron Slater determinant can also be written in compact notation as,

$$\Phi^{\mathrm{SD}} = \frac{1}{\sqrt{N!}} |\chi_i \chi_j \chi_k \cdots \chi_N\rangle \tag{4.18}$$

A Slater determinant incorporates exchange correlation effect which represents the correlation between the motion of two electrons with parallel spins.[308]

For the full electronic Hamiltonian H, the best wave function of the form defined in Eq. 4.18, is the one which minimize the electronic energy E_0 according to the variational principle,

$$E_0 = < \Phi_0 | H | \Phi_0 > \tag{4.19}$$

By minimizing E_0 with respect to chosen spin orbitals, the optimal set of spin orbitals can be determined that satisfy the Hartree-Fock equation of the form,

$$f_i \chi(\mathbf{x}_i) = \varepsilon \chi(\mathbf{x}_i) \tag{4.20}$$

where f_i is an effective one-electron Fock operator given by,

$$f_i = -\frac{1}{2} {\nabla_i}^2 - \sum_k^M \frac{Z_k}{r_{ik}} + v_i^{\mathrm{HF}} \tag{4.21}$$

The one-electron Hartree-Fock potential v_i^{HF} is the average potential experienced by the ith electron due to the presence of remaining $N-1$ electrons and thus depends on the spin orbitals of the other electrons and it is given by,

$$v_i^{\mathrm{HF}} = \sum_a J_a(\mathbf{x}_i) - K_a(\mathbf{x}_i) \tag{4.22}$$

where $J_a(\mathbf{x}_i)$ and $K_a(\mathbf{x}_i)$ represent a Coulomb operator and exchange operator respectively.

The Hartree-Fock equation (Eq. 4.20) is non-linear and must be solved iteratively following the self-consistent field (SCF) procedure.[308–310] For evaluation

of the matrix elements of the Hamiltonian H, the N-electron wave function is expanded as a linear combination of atomic centered basis functions (LCAO).

Although the Hartree-Fock based molecular orbital theory (HF-MO) is mathematically elegant, it suffers from many practical limitations[308–310] such as (i) choice of basis set poses a large problem of since the use of LCAO approach involves evaluation of large number of cumbersome two-electron integrals that appear in the Fock matrix elements and poses an upper limit for its application to large molecular systems; (ii) the Fock operator is an one-electron operator and except exchange, all electron correlation is ignored which are important in many molecular systems.

To circumvent these issues, two major methodological approaches have emerged over the years. These approaches are briefly summarized in the following sections.

4.2.3 Post Hartree-Fock approaches

To understand the behavior and properties of multi-electron systems, perhaps the most ideal approach is to solve the Schrödinger equation to obtain molecular energy levels and orbitals. However since the exact solution of the Schrödinger equation is almost impossible to obtain for most of the molecular systems, an alternate approach has evolved over the years to obtain the solution and is popularly known as post Hartree-Fock methods.[308]

Despite the limitations of the HF theory, the developments of post HF methods are motivated from the basics of the HF formalism. Thus the development of more and more rigorous techniques of post HF approach aim to reach the accuracy of the solutions of the HF equations in the infinite basis set i.e., to solve the HF equations without any additional approximations.

Since the calculations of HF equation with an infinite basis set is impractical, significant studies have been dedicated to construct wave functions to efficiently reach HF limit as intimately as possible. Several possibilities for constructing molecular orbitals from basis functions have been suggested including Slater-type orbitals (STOs),[311] Gaussian-type orbitals (GTOs),[312,313] and correlation-consistent polarized Core and Valence n-multiple Zeta functions (cc-pVnZ).[314,315]

The advantage and disadvantages of post HF based theories depend upon their ability to estimate the electron-correlation energy which is the difference between the true energy of the system and the energy of the system in the HF limit.[308,310,316,317]

4.2.4 Configuration Interaction (CI)

In Hartree-Fock theory, the electron-electron interaction is calculated in an average way such that each electron moves in the static field potential constructed from all other electrons. The so called dynamical correlation[318] which arises from the correlated motion of each electron with every other electron with different spin is thus ignored in HF theory (for reviews of electron correlation see refs 308, 317, 319). Although HF theory with sufficiently large basis set can account for most of the total energy, the remaining unaccounted difference in the energy known as correlation energy is often important for many chemical reactions.[310,317]

In Configuration Interaction (CI) methods, this problem is solved by constructing a multi-Slater determinant trial wave function of the form,

$$\Phi = a_0\Phi_{\mathrm{HF}}^{\mathrm{SD}} + \sum_i a_i\Phi_i^{\mathrm{SD}} \tag{4.23}$$

where Φ_i^{SD} represents the Slater-determinant wavefunctions.

The CI space thus represents the *N*-electron function space which is formed from the above defined *N*-electron basis functions. To determine the CI space, usually first a HF calculation is performed to calculate a set of MOs.

Currently available CI methods mainly differ in the approach of how the coefficients are determined and how the CI space is truncated. In the full CI scheme, all possible determinants are included in the CI space. Within a given one-electron basis set, full CI gives exact solution of the non-relativistic Schrödinger equation. However its applicability is limited to very small molecules due to significantly high computational requirements. To circumvent this problem, the number of excited determinants in the CI space is restricted. CI methods also give the energies of the excited states.[308] Several approaches have been proposed for this procedure (for reviews see refs 308, 317). In the current work, spectroscopy oriented configuration interaction (SORCI)[320] approach was used to calculate the excitation energies.

4.2.5 Spectroscopically Oriented Configuration Interaction

In SORCI, the first-order interacting space is divided into strong and weak perturbing configurations. The strong configurations are treated variationally while the weak configurations are treated with second-order Møller-Plesset Perturbation

Theory (for details of the method see ref 308). The use of approximate natural orbitals eliminates the problem of choosing a suitable single-particle basis whose quality would affect the final CI result.

SORCI gains computational efficiency by use of several thresholds which have been carefully adjusted for the system under study. The thresholds used are: $T_{Pre} = 10^{-3}$, $T_{Nat} = 10^{-6}$ and $T_{Sel} = 10^{-6}E_h$ (see supporting information of ref 321). Only core orbitals are frozen. Ahlrich's SV(P) basis set[322] was used which is appropriate for the calculations of the complete chromophore. SORCI has been successfully used to calculate vertical excitation energies for retinal proteins previously.[321]

4.2.6 Semiempirical Approaches to the HF-MO theory

As described in section 4.2.2, the computational efficiency of HF calculations is strongly influenced by a requisite to evaluate the two-electron integrals.[308] The popularity and efficiency of semi-empirical MO theories originate from their approach to accurately estimate or in some cases neglect these computationally expensive integrals rather than their explicit evaluation and effectively calculate the matrix elements of the HF secular equation. Over the years several approaches have been developed. These approaches are reviewed extensively in many books[323–327] and articles.[328–331]

In general, available semi-empirical methods employ a minimal basis set for the valence electrons and thus ignore all core electrons. Depending upon the treatment of integral approximation, current semi-empirical methods can be broadly categorized into three levels: (i) complete neglect of differential overlap (CNDO);[332,333] (ii) intermediate neglect of differential overlap (INDO);[334–336] and (iii) neglect of diatomic differential overlap (NDDO). Most popular current semi-empirical methods such as MNDO,[337,338] AM1[339] and PM3[340] are based on the NDDO approach and differ in the different treatment of nuclear core-core repulsion. Each method uses only *s*- and *p*-functions, which are taken as STOs. In the NDDO approach, only the two-center two-electron integrals with atomic basis functions located on same atomic centers are considered.

4.3 Density Functional Theory

An alternative approach to solve the Schrödinger equation is to use the electron density $\rho(\mathbf{r})$ to form the Hamiltonian of the system where the number of electrons N correspond to integration of $\rho(\mathbf{r})$ over all space,

$$N = \int \rho(\mathbf{r}) d\mathbf{r} \tag{4.24}$$

With this spirit, Thomas and Fermi in 1927 calculated kinetic energy of a uniform electron gas as a function of the electron density and thus represented kinetic and also potential energy terms as density functionals. This development is perhaps one of the most early efforts to formulate a density function theory (DFT).[341] However DFT was established as a robust quantum mechanical method only after the seminal work[342] of Hohenberg and Kohn in 1964.

4.3.1 Hohenberg-Kohn Theorems

Hohenberg and Kohn proposed two theorems to establish the use of electron density $\rho(\mathbf{r})$ as a fundamental variable of the system and provided rigorous foundation for DFT. According to the *existence theorem*,[342] the ground state electron density determines the external potential. Thus the Hamiltonian, the wave function, the energy and all other properties are determined by the electron density. The total energy is given by,

$$E_0 = E[\rho(r)] \tag{4.25}$$

The second Hohenberg-Kohn theorem is known as the *variational theorem* which states that for a chosen well-behaved density, the energy expectation value (E_{ch}) of the system must be greater than or equal to the true ground state energy (E_0),

$$E_0 \leq E_{\text{ch}} \tag{4.26}$$

However, the theorem did not provide any procedure to obtain the ground-state density of a system in practice. The developments of DFT methods are hence focused towards finding a functional that connects the electron density with the energy.[341,343]

4.3.2 Kohn-Sham Theory

Although the Hohenberg-Kohn theorems established that the electron density determines the system Hamiltonian, wave function and energy; this alternative approach was as complex as the MO theory due to the presence of the electron-electron interaction term in the Hamiltonian. Kohn and Sham first proposed that the Hamiltonian of the system can be expressed as a sum of one-electron operators. The eigenfunctions of such a Hamiltonian are Slater determinants of the individual one-electron eigenfunctions, and the eigenvalues as the sum of the one-electron eigenvalues.[344] The charge density $\rho(\mathbf{r})$ is expressed in terms of one-particle wave functions ϕ_i as,

$$\rho(\mathbf{r}) = \sum_{i}^{N} |\phi_i|^2 \tag{4.27}$$

The formulation of Kohn-Sham (KS) theory assumed that a system of non-interacting electrons can be chosen such that their overall ground state density is same as that of a system with interacting electrons. The energy functional for a system with N electrons and M nuclei as a function of charge density $\rho(\mathbf{r})$ is written as,

$$\begin{aligned} E[\rho(\mathbf{r})] &= \sum_{i}^{N} \left\langle \phi_i | -\frac{1}{2}\nabla_i^2 | \phi_i \right\rangle - \sum_{i}^{N} \left\langle \phi_i | \sum_{k}^{M} \frac{Z_k}{|\mathbf{r}_i - \mathbf{r}_k|} | \phi_i \right\rangle \\ &+ \sum_{i}^{N} \left\langle \phi_i | \frac{1}{2} \int \frac{\rho(\mathbf{r}')}{|\mathbf{r}_i - \mathbf{r}'|} d\mathbf{r}' | \phi_i \right\rangle + E_{\mathrm{xc}}[\rho(\mathbf{r})] \\ &+ \frac{1}{2} \sum_{k,l}^{M} \frac{Z_k Z_l}{|\mathbf{r}_k - \mathbf{r}_l|} \end{aligned} \tag{4.28}$$

where ϕ_i are the Kohn-Sham orbitals and the terms on the r.h.s. refer to the kinetic energy of the non-interacting electrons, classical nuclear-electron attraction, classical electron-electron repulsion and exchange-correlation energy respectively. The exchange-correlation energy E_{xc} represents the summation of the correction to the kinetic energy arising from the interacting electrons, and non-classical corrections to the electron-electron repulsion energy.

The application of the variational principle gives the orbitals ϕ_i that minimize the energy E in Eq. (4.28), that are the eigenfunctions of KS one-electron Hamil-

tonian operator h^{KS} is defined as,

$$h_i^{\mathrm{KS}} = -\frac{1}{2}\nabla_i^2 - \sum_k^M \frac{Z_k}{|\mathbf{r}_i - \mathbf{r}_k|} + \int \frac{\rho(\mathbf{r}')}{|\mathbf{r}_i - \mathbf{r}'|} d\mathbf{r}' + \frac{\delta E_{\mathrm{xc}}}{\delta \rho} \tag{4.29}$$

The last term on the r.h.s of Eq. 4.29 is referred to as the functional derivative of the exchange-correlation energy E_{xc} i.e.,

$$V_{\mathrm{xc}} = \frac{\delta E_{\mathrm{xc}}}{\delta \rho} \tag{4.30}$$

The orbitals ϕ_i satisfy an eigenvalue problem with eigenvalues ε_i,

$$h_i^{\mathrm{KS}} \phi_i = \varepsilon_i \phi_i \quad \text{where} \quad i = 1, \ldots, N \tag{4.31}$$

The KS Hamiltonian itself depends on the density which is determined from the orbitals as in Eq. 4.27, the Kohn-Sham eigenvalue equation Eq. 4.31 must be solved self-consistently.

In the limit of an infinite basis set, KS-DFT is an exact theory in the sense that no approximations are considered during the formulation.[310,341] This exactness of the theory accurately estimates the electron correlation energies and hence has seen wide applications over the years. Increasing the size of basis set allows better and better representation of KS orbitals, thus increasing the computational cost. This puts limitations on DFT's applicability to large systems such as biomolecules.

4.3.3 The self-consistent-charge density-functional tight-binding method (SCC-DFTB)

Although more efficient than HF theory, the exact solution of DFT for large molecular systems is still computationally very demanding.[317] Consequently various alternative formulations have been developed to increase DFT efficiency. One such method is the self-consistent-charge density-functional tight-binding method (SCC-DFTB)[345] which can be considered as an extension of the non-self-consistent density functional tight-binding (DFTB) method.[346]

In SCC-DFTB, the DFT total energy functional is written as a functional of

charge density $\rho(\mathbf{r})$ is given as,

$$\begin{aligned} E &= \sum_{i}^{\text{occ.}} \left\langle \phi_i \left| -\frac{1}{2}\nabla_i^2 + \frac{1}{2}\int \frac{\rho(\mathbf{r}')}{|\mathbf{r}_i - \mathbf{r}'|} d\mathbf{r}' - \sum_{k}^{M} \frac{Z_k}{|\mathbf{r}_i - \mathbf{r}_k|} \right| \phi_i \right\rangle + E_{\text{xc}}[\rho(\mathbf{r})] \\ &+ \frac{1}{2}\sum_{k,l}^{M} \frac{Z_k Z_l}{|\mathbf{r}_k - \mathbf{r}_l|} \end{aligned} \tag{4.32}$$

where the last term on the r.h.s. represents the nuclear-nuclear repulsion between k^{th} and l^{th} nuclei.

The charge density $\rho(\mathbf{r})$ can be written as a superposition of a reference or input density $\rho_0(\mathbf{r})$ and a small fluctuation $\delta\rho(\mathbf{r})$.[347] The total energy is then expanded around a reference density $\rho_0(\mathbf{r})$ upto second order in the density fluctuations $\delta\rho(\mathbf{r})$. Without the expansion of E_{xc}, the total energy is thus written as,[345]

$$\begin{aligned} E &= \sum_{i}^{\text{occ.}} \left\langle \phi_i \left| -\frac{1}{2}\nabla_i^2 + \frac{1}{2}\int \frac{\rho_0(\mathbf{r}')}{|\mathbf{r}_i - \mathbf{r}'|} d\mathbf{r}' - \sum_{k}^{M} \frac{Z_k}{|\mathbf{r}_i - \mathbf{r}_k|} + V_{\text{xc}}[\rho_0(\mathbf{r})] \right| \phi_i \right\rangle \\ &- \frac{1}{2}\int\int \frac{\rho_0(\mathbf{r}')\left(\rho_0(\mathbf{r}) + \delta\rho(\mathbf{r})\right)}{|\mathbf{r}_i - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}' - \int V_{xc}[\rho_0(\mathbf{r})]\left(\rho_0(\mathbf{r}) + \delta\rho(\mathbf{r})\right) d\mathbf{r} \\ &+ \frac{1}{2}\int\int \frac{\delta\rho_0(\mathbf{r}')\left(\rho_0(\mathbf{r}) + \delta\rho(\mathbf{r})\right)}{|\mathbf{r}_i - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}' + E_{\text{xc}}[\rho_0(\mathbf{r}) + \delta\rho(\mathbf{r})] \\ &+ \frac{1}{2}\sum_{k,l}^{M} \frac{Z_k Z_l}{|\mathbf{r}_k - \mathbf{r}_l|} \end{aligned} \tag{4.33}$$

Subsequently after substitution of the Taylor-expansion of the exchange-correlation E_{xc} at the reference density and after cancellation of terms that are linear in $\delta\rho(\mathbf{r})$, the total energy becomes,

$$\begin{aligned} E &= \sum_{i}^{\text{occ.}} \left\langle \phi_i \left| \hat{H}_0[\rho_0(\mathbf{r})] \right| \phi_i \right\rangle \\ &- \frac{1}{2}\int\int \frac{\rho_0(\mathbf{r}')\rho_0(\mathbf{r})}{|\mathbf{r}_i - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}' - \int V_{\text{xc}}[\rho_0(\mathbf{r})]\rho_0(\mathbf{r}) d\mathbf{r} + E_{\text{xc}}[\rho_0(\mathbf{r})] + \frac{1}{2}\sum_{k,l}^{M} \frac{Z_k Z_l}{|\mathbf{r}_k - \mathbf{r}_l|} \\ &+ \underbrace{\frac{1}{2}\int\int \left(\frac{1}{|\mathbf{r}_i - \mathbf{r}'|} + \left.\frac{\delta^2 E_{\text{xc}}}{\delta\rho(\mathbf{r})\delta\rho(\mathbf{r}')}\right|_{\rho_0(\mathbf{r})} \right) \delta\rho(\mathbf{r})\delta\rho(\mathbf{r}') d\mathbf{r} d\mathbf{r}'}_{E_{2\text{nd}}} \end{aligned} \tag{4.34}$$

where $\hat{H}_0[\rho_0(\mathbf{r})]$ represents the KS Hamiltonian operator resulting from a reference density $\rho_0(\mathbf{r})$ from Eq. 4.29. The second, third and fourth term on the r.h.s.

of Eq. 4.34 together represent the energy contributions, E_{rep}. As in standard TB approach the contributions represented by E_{rep} are pairwise, repulsive and short-range,[347] while the last term on the r.h.s. of Eq. 4.34 represents the second-order contributions E_{2nd} in the density fluctuations $\delta\rho(\mathbf{r})$.

Eq. 4.34 can be thus rewritten as,

$$E = \sum_{i}^{\text{occ.}} \left\langle \phi_i \left| \hat{H}_0[\rho_0(\mathbf{r})] \right| \phi_i \right\rangle + E_{\text{rep}}[\rho_0(\mathbf{r})] + E_{\text{2nd}} \tag{4.35}$$

The standard non-SCC DFTB approach in which the second-order correction E_{2nd} is neglected, gives reasonable results for chemical reactions where charge transfer can be neglected. Thus, the zeroth-order energy functional is given as,

$$E_0^{\text{TB}} = \sum_{i}^{\text{occ.}} \left\langle \phi_i \left| \hat{H}_0[\rho_0(\mathbf{r})] \right| \phi_i \right\rangle + E_{\text{rep}}[\rho_0(\mathbf{r})] \tag{4.36}$$

For the solution of Kohn-Sham equations, the single-particle wave functions ϕ_i within an LCAO approach are expanded into a suitable set of localized atomic orbitals η_ν,

$$\phi_i(\mathbf{r}) = \sum_{\nu} c_{\nu i} \eta_\nu (\mathbf{r} - \mathbf{r}_k) \tag{4.37}$$

These atomic orbitals are chosen in a Slater-type representation,[346] which are determined by solving a modified Schrödinger equation for a free neutral pseudoatom with self-consistent DFT calculations with a minimal basis set.[345] Further the reference density $\rho_0(\mathbf{r})$ is written as the sum over atomic reference densities,

$$\rho_0(\mathbf{r}) = \sum_{k}^{M} \rho_0^k(\mathbf{r}) \tag{4.38}$$

Following the variational principle to the energy functional of Eq. (4.36), the non-SCF KS equations are rearranged as,

$$\sum_{\nu}^{N} c_{\nu i} \left\langle \eta_\mu | \hat{H}_0 | \eta_\nu \right\rangle = \sum_{\nu}^{N} c_{\nu i} \varepsilon_i \left\langle \eta_\mu | \eta_\nu \right\rangle \quad \forall \mu, i \tag{4.39}$$

The matrix elements of $H_{\mu\nu} = \left\langle \eta_\mu | \hat{H}_0 | \eta_\nu \right\rangle$ are evaluated by introducing many approximation as,[345]

$$H^0_{\mu\nu} = \begin{cases} \varepsilon_\mu^{\text{neutral free atom}} & \text{if } \mu = \nu \\ \left\langle \eta_\mu^k | \hat{T} + \hat{V}[\rho_0^k(\mathbf{r}) + \rho_0^l(\mathbf{r})] | \eta_\nu^l \right\rangle & \text{if k} \neq 1 \\ 0 & \text{otherwise.} \end{cases} \tag{4.40}$$

Following a two center approximation, only two-center Hamiltonian matrix elements are treated and explicitly evaluated in combination with the two-center overlap matrix elements given by,

$$S_{\mu\nu} = \langle \eta_\mu | \eta_\nu \rangle \tag{4.41}$$

The matrix elements of $H^0_{\mu\nu}$ and the overlap matrix $S_{\mu\nu}$ are tabulated for a range of nuclear distances.

Second Order Corrections

The above DFTB formalism gives appropriate results for structures where the electron density can be well represented as a sum of atomic-like densities. However for simulation of chemical reactions that involve significant charge transfer the earlier neglected $E_{2\text{nd}}$ term is necessary for estimation of correct charge transfer. This extended approach is known as the self-consistent charge DFTB (SCC-DFTB).[345] The last term on the r.h.s. of Eq. (4.34)) is written as,

$$E_{2\text{nd}} = \frac{1}{2}\sum_{k,l}^{M}\frac{1}{2}\int\int\left(\frac{1}{|\mathbf{r}-\mathbf{r}'|} + \frac{\delta^2 E_{\text{xc}}}{\delta\rho(\mathbf{r})\delta\rho(\mathbf{r}')}\bigg|_{\rho_0(\mathbf{r})}\right)\delta\rho_k(\mathbf{r})\delta\rho_l(\mathbf{r}')d\mathbf{r}d\mathbf{r}' \tag{4.42}$$

The density fluctuations $\delta\rho(\mathbf{r})$ are decomposed into atomic contributions which are expanded into series of radial and angular functions. Thus, $\delta\rho^k(\mathbf{r})$ term can be written as,

$$\delta\rho^k(\mathbf{r}) = \sum_{a,b} K_{ba} F^k_{ba}(\mathbf{r}-\mathbf{r}_k) Y_{ab}\left(\frac{\mathbf{r}-\mathbf{r}_k}{|\mathbf{r}-\mathbf{r}_k|}\right) \tag{4.43}$$

where $F^k_{ba}(\mathbf{r})$ denotes the normalized radial dependence of the density fluctuations on atom m for the corresponding angular momentum. The most important contributions from the charge transfers between different atom are accounted by the monopole term, while the higher-order terms decay rapidly with increasing the interatomic distance. Thus truncating after the monopole term, Eq. 4.43 becomes,

$$\delta\rho^k(\mathbf{r}) \approx \Delta q_k F^k_{00}(|\mathbf{r}-\mathbf{r}_k|) Y_{00} \tag{4.44}$$

Note that, the total charge of the system is thus given by sum of charge fluctuations (Δq) of all atoms,

$$\int \delta\rho(\mathbf{r}) = \sum_k \Delta q_k \quad \text{where} \quad \Delta q_k = q_k - q^0_k \tag{4.45}$$

where q_k stands for the Mulliken charge and q_k^0 represents the number of valence electrons of an neutral atom.

The second-order energy term can now be written as,

$$E_{\text{2nd}} = \frac{1}{2}\sum_{k,l}^{M} \Delta q_k \Delta q_l \gamma_{kl} \tag{4.46}$$

where γ_{mn} denotes,

$$\gamma_{kl} = \frac{1}{4\pi}\int\int \left(\frac{1}{|\mathbf{r}-\mathbf{r}'|} + \frac{\delta^2 E_{xc}}{\delta\rho(\mathbf{r})\delta\rho(\mathbf{r}')}\bigg|_{\rho_0(\mathbf{r})} \right) F_{00}^{k}(|\mathbf{r}-\mathbf{r}_k|) F_{00}^{l}(|\mathbf{r}-\mathbf{r}_l|) d\mathbf{r} d\mathbf{r}' \tag{4.47}$$

For cases where the charges are located at one and the same atom .i.e. for $k = l$, γ_{kk} can be approximated by the difference of the atomic ionization potential and the electron affinity represented by the Hubbard parameter U_k,[345]

$$\gamma_{kk} \approx 2\xi_k \approx U_k \tag{4.48}$$

where ξ_k represents the chemical hardness of atom k.[345] On the other hand, for cases where $k \neq l$, γ_{kl} is determined analytically and depends on the extension of the charge densities of atoms k and l and thus represented as,[345]

$$\gamma_{kl} = \frac{1}{|\mathbf{r}_k - \mathbf{r}_l|} - S \tag{4.49}$$

where S denotes a short-range function so as to give desired results for small interatomic distances.[345]

Total energy and DFTB secular equation

Substituting all above approximations, the DFT total energy upto second-order is written as,

$$E_{\text{SCC-DFTB}} = \sum_{i}^{\text{occ.}} \left\langle \phi_i \left| \hat{H}_0[\rho_0(\mathbf{r})] \right| \phi_i \right\rangle + \frac{1}{2}\sum_{k,l}^{M} \Delta q_k \Delta q_l \gamma_{kl} + E_{\text{rep}}[\rho_0(\mathbf{r})] \tag{4.50}$$

The atomic charges depend on the single particle wave functions ϕ_i which requires a self-consistent procedure to minimize the energy functional of Eq. (4.50). The wave functions ϕ_i are expanded using LCAO with expansion coefficients $c_{\nu i}$.

The solution of the Kohn-Sham equations are obtained by applying the variational principle that minimize the energy in Eq. (4.50). Thus, the resulting Kohn-Sham secular equations are written as,

$$\sum_{\nu}^{M} c_{\nu i}\left(H_{\mu\nu} - \varepsilon_i S_{\mu\nu}\right) = 0 \;\; \forall \;\; \mu, i \tag{4.51}$$

The method predicts accurate geometries for covalently bonded systems and has been applied to many biological systems.[348]

Extension of the SCC-DFTB Method

The SCC-DFTB second-order formalism assumes a fixed chemical hardness parameter independent of the atomic charge state. Further, the effective charge-charge interactions depend on the inverse relation of the atomic size and chemical hardness. Nevertheless, for better treatment of hydrogen bonding interactions and anionic molecules with localized net charges, the SCC-DFTB method was extended to incorporate the third-order terms in the charge density fluctuations and to improve the interatomic electrostatic contributions[349]

For X-H pairs (where X denotes heavy atoms and H denotes hydrogen), γ_{kl} is modified (γ_{kl}^{mod}) which results in appropriate faster decay by including additional damping term,

$$\gamma_{kH}^{\text{mod}} = \frac{1}{|\mathbf{r}_k - \mathbf{r}_H|} - S \exp\left[-\left(\frac{U_k + U_H}{2}\right)^{\zeta} R_{kH}^2\right] \tag{4.52}$$

where the exponent ζ can be fitted based on binding energies and proton affinities of numerous gas-phase molecules.[349]

To include the charge dependence of the Hubbard parameter (U_k), the DFT total energy is expanded up to third order in the density fluctuations $\delta\rho$ (for clarity, $\delta\rho(\mathbf{r}) \equiv \delta\rho$ is assumed),

$$\begin{aligned} E[\rho] = E[\rho_0] \;\; &+ \;\; \int \left[\frac{\delta E[\rho]}{\delta\rho}\right]_{\rho_0} \delta\rho\, d\mathbf{r} \\ &+ \;\; \frac{1}{2}\int\int \left[\frac{\delta^2 E[\rho]}{\delta\rho\,\delta\rho'}\right]_{\rho_0} \delta\rho\,\delta\rho' d\mathbf{r} d\mathbf{r}' \\ &+ \;\; \frac{1}{6}\int\int\int \left[\frac{\delta^3 E[\rho]}{\delta\rho\,\delta\rho'\delta\rho''}\right]_{\rho_0} \delta\rho\,\delta\rho'\delta\rho'' d\mathbf{r} d\mathbf{r}' d\mathbf{r}'' \end{aligned} \tag{4.53}$$

Following approximations, the third-order term depends on the derivative of the Hubbard parameter (U_k) with respect to the atomic charge q_k denoted as (U_k^d),

$$E_{kkk}^{3\text{rd}} \approx \frac{1}{6} U_k^d \Delta q_k^3 \tag{4.54}$$

It has been shown that with these modifications SCC-DFTB can accurately describe[349] the proton affinities and hydrogen bonding interactions and hence has been employed in the current work.

4.4 Molecular Mechanics Force Field

In principle to understand molecular properties, quantum mechanical methods should be employed especially for properties where the electronic degrees of freedom play significant role. Although these methods can provide accurate results, the simulations in most cases are very time consuming. Hence to understand the properties of systems which do not depend on the electronic distribution in a molecule, alternative approaches such as molecular mechanics (also known as force field methods), Brownian dynamics, and fluid dynamics have been developed over the years.

The force field methods include electronic motions only implicitly and follow the Born-Oppenheimer approximation. Thus the energy of the molecular system is calculated as a parametric function nuclear positions following Newtonian dynamics. Further, the atoms are described by point-like centers with associated mass and partial charges. The interactions between these atom points are described with contributions from changes in the bond length, the rotation around bonds, and changes in the angles and interactions between non-bonded atoms.

The motivation of the development of force field methods arises from the observed fact that the subunits of a molecular system tend to be structurally similar in different molecules. Most of the force field methods discriminate different atoms in various chemical environments by categorizing them into atom types. Thus, an atom type contains information not only about the element of an atom, but also about its hybridization state and the local chemical environment.

The aim of development of force field methods is not only to reproduce the structural properties, but also to predict them. Hence over the years, numerous force field methods have been designed to predict particular structural properties of a system for instance, to simulate protein folding process. Thus there is no unique

way to design a particular force field and in this sense the force field methods are considered empirical.

Transferability and accuracy are two essential issues which can be used to categorize available force field methods. Transferability refers to the ability of a force field to model a variety of molecular systems. Some examples of such force fields include Universal force field (UFF)[350] or the MM2, MM3, and MM4 force fields developed by Allinger and co-workers[351–356] Numerous other force fields have been developed to predict molecular properties as accurately as possible for example, CHARMM,[357] AMBER,[358,359] OPLS[360] and GROMOS.[361]

4.4.1 The CHARMM Force Field

The chemistry at Harvard macromolecular mechanics (CHARMM)[357] force fields include a set of force fields for simulating proteins, lipids and nucleic acids. For protein simulations, united atom CHARMM19 force field[362] and all-atom CHARMM22 force field[363] have been developed. For the current work, all-atom CHARMM force field was employed. While the form of the energy function in CHARMM is similar to other force fields such as AMBER[358,359] or GROMOS,[361] distinct differences exist in the determination of the force field parameters.[363]

The CHARMM energy function contains terms arising from both internal interactions due to bonds and external interactions due to non-bonded interactions. The force field parameterization is based on a wide range of experimental and *ab initio* data.[363] In the following section, the individual energy terms are briefly summarized.

Internal Interaction Terms

The internal interaction terms represent the interactions of the bonded atoms. These are modelled using the internal coordinate system in which the position of each atom is specified relative to other atoms of the molecule and the changes in the molecular geometry are described by changes in bond lengths, angles, and dihedral angles of the molecule (Figure 4.1) and not by the changes in its cartesian coordinates.

The change in the bond length (Figure 4.1A) is usually represented by a harmonic potential using Hooke's law which states that the variation in the energy is

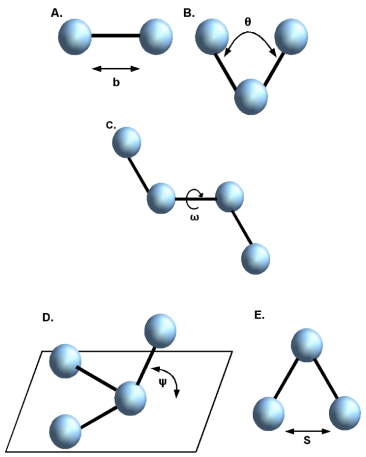


Figure 4.1: Internal interaction terms used in the force field energy function are based on the changes in the geometries of: A. bond lengths, B. bond angles, C. dihedral angles, D. improper angles, and E. Urey-Bradley.

proportional to the square of the change in the bond length from the equilibrium bond length b_0,

$$E_{\text{bond}} = \frac{1}{2} K_{\text{bond}} (b - b_0)^2 \tag{4.55}$$

where K_{bond} is the associated force constant.

The Harmonic approximation gives sensible results when the change in the bond length is small relative to the equilibrium bond length. However, to accurately model large deviations more complicated functional forms have to be considered. One possible extension is to include cubic and higher contributions to the harmonic potential as done in MM2 force field. Another possibility is the Morse potential[364] which accurately predicts the asymptotic nature of bond deformation. However the Morse potential requires three parameters to be specified for each bond and thus compromises the computational efficiency.

In addition to represent the bond length deformations, the harmonic potentials are also used to represent the variation of angles (Figure 4.1B) from their equilibrium values θ_0 with the associated force constant K_{angle}. Similar to the bond-stretching terms, the accuracy of the force field can be improved by including the higher-order terms to the harmonic potential for angles.

In a molecular simulation, to cause large change in the bond lengths and angles from their equilibrium values, substantial energy is required and therefore the vari-

ation in structure and relative energies in a simulation is due to the contributions from the torsional and non-bonded terms. The torsional energy contribution is associated with energy required for the rotation around the central bond in a quartet of atoms (Figure 4.1C). Thus to properly account the energy contribution arising from the changes in the dihedral angles ω, the torsional potential requires to be periodic in ω. One such functional form is,

$$E_{\text{torsion}} = \sum_{\text{torsion}} K_{\text{torsion}} \left(1 + \cos(n\omega - \delta_n)\right) \tag{4.56}$$

where n and K_{torsion} are different periodicities and force constant respectively.

In many molecular systems, all atoms of a quartet do not lie in the same plane (Figure 4.1D) and an additional force field term is necessary to model such deviation from planarity. Such out-of-plane contributions are often modelled using a harmonic potential for the angle ψ between a bond from the central atom and the plane defined by the central atom and the other two atoms (Figure 4.1D).

To properly account the vibrational properties for cases such as H-C-H angles, CHARMM force field contains an additional term known as the Urey-Bradley term (E_{UB}). This contribution is represented as a harmonic potential for the distance S (Figure 4.1E) between the outer atoms of a triple of atoms connected by an angle term,

$$E_{\text{UB}} = \frac{1}{2} K_{\text{UB}} (S - S_0)^2 \tag{4.57}$$

where K_{UB} and S_0 are the force constant and equilibrium distance respectively.

External Interaction Terms

The external interaction terms or the non-bonded interaction terms include electrostatic and van der Waals interactions which are mediated through space and are independent of a specific bonding relationship between atoms. These interactions are usually modelled as function of some inverse power of the distance.

The electrostatic interactions are usually represented by using Coulomb's law where the interactions between partial atomic charges q_i associated with the nuclear centers are separated by distance r_{ij} in space,

$$E_{\text{elec}} = \frac{1}{4\pi\varepsilon_0} \sum_{i<j} \frac{q_i q_j}{r_{ij}} \tag{4.58}$$

where i and j correspond to the non-bonded atoms and ε_0 is the permittivity of vacuum.

Several other more elaborate approaches have also been developed over the years to accurately treat the electrostatic interactions and are described below.

As in the case of rare gas atoms, the electrostatic interactions can not account for all of the non-bonded interactions. The deviations from ideal gas behavior of such cases were first studied by van der Waals and hence these contributions are known as van der Waals contributions. Van der Waals contributions have two components: long-range attractive contribution and short-range repulsive contribution.

The attractive contribution is due to the dispersive force that arises because of instantaneous dipoles formed as a result of the fluctuations in the electron clouds. In a molecule, an instantaneous dipole can consequently induce a dipole in neighboring atoms giving rise to an attractive inductive effect. The dispersion force is sometimes known as London force.[365] The dispersive interactions may be represented using Drude model which consists of molecules with two charges $+q$ and $-q$ separated by distance r.[366]

The short-range repulsive contribution also has a quantum mechanical origin and arises due to electrons with same spin and therefore the repulsive force is sometimes referred to as exchange force or overlap force. The effect of exchange is to reduce the electrostatic repulsion between pairs of electrons by prohibiting them from occupying the same region of space.

The van der Waals interactions are usually represented as the Lennard-Jones 12-6 function,

$$E_{\mathrm{vdW}} = \sum_{i<j} 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] \tag{4.59}$$

where ε is the collision parameter that represents the separation for which the energy is zero, and σ represents the well-depth where the interaction energy is minimum. The Lennard-Jones potential has an attractive interaction that varies as r^{-6} and a repulsive interaction that varies as r^{-12}. Several other alternatives have also been proposed to model the van der Waals interactions.

To summarize, the complete CHARMM energy function with all the internal

and external terms described earlier,

$$\begin{aligned} E &= \sum_{\text{bonds}} \frac{1}{2} K_{\text{bonds}} (b - b_0)^2 + \sum_{\text{angles}} \frac{1}{2} K_{\text{angles}} (\theta - \theta_0)^2 \\ &+ \sum_{\text{torsion}} K_{\text{torsion}} (1 + \cos(n\omega - \delta_n)) + \sum_{\text{impropers}} \frac{1}{2} K_{\text{imp}} (\psi - \psi_0)^2 + \frac{1}{2} K_{\text{UB}} (S - S_0)^2 \\ &+ \frac{1}{4\pi\varepsilon_0} \sum_{i<j} \frac{q_i q_j}{r_{ij}} + \sum_{i<j} 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] \end{aligned} \tag{4.60}$$

Following the representation of the energy function, it is necessary to parameterize various force constants and associated equilibrium parameters for bonds, angles, dihedrals, impropers and Urey-Bradley terms. The parameterization ensures wider applicability of the force field to a variety of systems.

In CHARMM the optimization of the parameters was carried out iteratively to achieve self-consistency among the different terms of the potential energy function.[363] The parameters for the external contributions (Coulomb and Lennard-Jones) were chosen from either earlier developed CHARMM parameter sets or based on the reproduction of *ab initio* interaction calculations on rigid monomers. Starting from these parameters, parameters of the internal interaction terms (bonds, angles, dihedrals, impropers and Urey-Bradley) were determined using structural and vibrational data for the model compounds.[363]

Chapter 5

Theoretical Methods II: Multi-Length-Scale Approaches

Summary

This chapter introduces multi-length-scale approaches that have been developed over the years and which are relevant to the current work. The methodological details of the hybrid quantum mechanics/molecular mechanics approach and its implementation in CHARMM are summarized in section 5.2. The importance of proper treatment of electrostatic interactions is outlined in section 5.3 followed by the description of the generalized solvent boundary potential (GSBP) method used in the work. The chapter concludes by summarizing the implementation of GSBP within the framework of SCC-DFTB/MM in CHARMM package.

5.1 Introduction

For understanding complex biological processes, it is essential to investigate several properties of the biological system. Although associated with the same molecular system, these processes may become dominant at different stages of the system and differ in time and/or system size. In this sense the biological processes can be considered as multi-length-scale processes.

Computer simulation techniques that have been developed over the years usually work better for certain type of processes. For instance, accurate but expensive *ab initio* methods can reveal details at electronic levels, but their applications are

limited at most to a few hundred atoms. Empirical force field methods can simulate conformational changes as a function of time for several thousands of atoms in a molecular system but can not investigate reactions which involve formation and breaking of chemical bonds. Brownian dynamics and hydrodynamics simulations can investigate very large molecular systems that are involved in processes at cellular or sub-cellular level such as protein folding or formation of aggregated emulsions, but usually model atoms as a part of molecular subunit or subunits.

One interesting example of a multi-length-scale biological process is the mechanism of PT in a membrane protein, bacteriorhodopsin (bR). To understand the isomerization of retinal in bacteriorhodopsin, quantum mechanical treatment is necessary, while to understand the large conformational changes of the protein during the photocycle, long time scale molecular dynamics is necessesary. Further, investigation of the mechanism of PT that takes place over several angstroms distance, requires quantum mechanical treatment of protein groups that could be involved in the PT pathway as well as suitable techniques to model the solvent effects of the surrounding lipid membrane environment and water molecules.

Consequently several techniques have been proposed to understand such multi-length-scale processes. These techniques decompose the overall system into subsystems based on the dominant process of the subsystem. Each subsystem is then treated with a method or a range of methods that are designed to accurately and efficiently predict specific properties. The results of these subsystems are combined to arrive at a coherent picture of the overall process. In the following subsections, multi-length-scale techniques employed in the current work are summarized.

5.2 Quantum Mechanics/Molecular Mechanics Approach

In many biomolecular systems, changes in the electronic structure are often limited to a small part of the whole system, while the rest of the system may be responsible for performing other important roles such as to undergo conformational changes and to provide structural stability to the system. The subsystem in which chemical reactions involve formation and breaking of bonds, can be treated with quantum mechanics (QM) methods, while the rest of the system can be treated with classical molecular mechanics (MM) methods (Figure 5.1). In a pioneering work, Warshel and Levitt[367] studied enzymatic reactions using a method that in-

corporated all aspects of the hybrid QM/MM method. Few years later, Singh and Kollman[368] carried out an *ab initio* QM/MM structure optimization using Hartree-Fock as QM method.

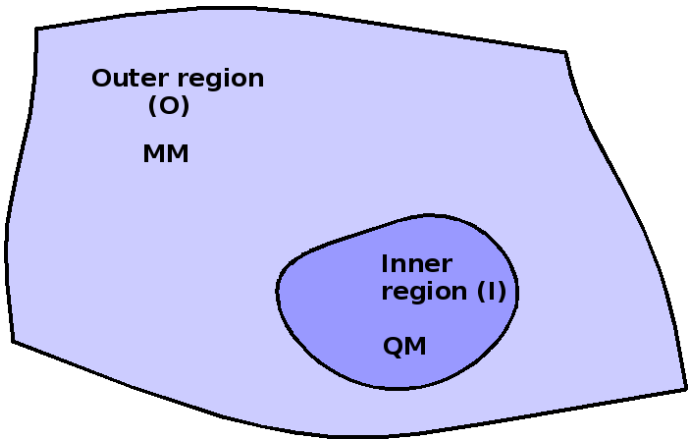


Figure 5.1: To simulate large systems that require explicit representation, in a hybrid QM/MM approach, the system (S) is divided into two subsystems. The inner region (I) in which bond-breaking/bond-making processes take place are treated with a QM method, and the remaining large outer region (O) which is treated with MM.

Since then, numerous studies have employed hybrid QM/MM methods to investigate complex biomolecular reactions and have illustrated the efficiency of the approach. Over the years, several variations of hybrid QM/MM approach have been proposed, making it a popular tool to study processes such as enzyme catalysis and to elucidate mechanism of ion transfers in biomolecules (for some recent reviews see refs 369–373).

In most cases since QM and MM regions strongly interact, it is difficult to express the total energy of the entire system simply as a sum of the energies of the subsystems. Therefore to properly describe the energy of the entire system, it is often necessary to introduce additional coupling terms and to take precautions to treat the boundary region. The boundary region is usually considered as the region where the QM and MM approaches are augmented or modified. The definition of boundary region is not unique and depends on the employed QM/MM approach.

5.2.1 QM/MM Boundary

One of the most sensitive issue in the QM/MM formalism is the description of QM/MM boundary region and consequently several approaches have been proposed for the treatment of the QM/MM boundary region.[369,370]

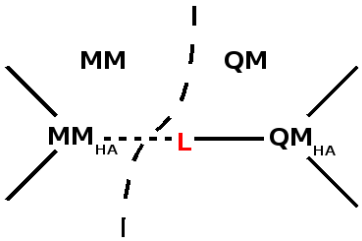


Figure 5.2: The boundary between QM and MM regions involve introduction of a QM link atom (L) across a covalent bond between the MM host atom (MM_{HA}) and the QM host atom (QM_{HA}). The charge neutral group that includes the MM_{HA} is termed as MM host group (MM_{HG}).

Perhaps the most straightforward approach is to define QM and MM regions such that the boundary does not pass through a covalent bond. Such approaches are employed widely for explicit-solvation systems or for enzymatic reactions in which the reactants such as substrates or cofactors are not covalently bound to the enzyme. However, in most cases it is impossible to avoid the QM/MM boundary cut across a covalent bond. In such cases, proper treatment is necessary to avoid dangling bonds in the frontier QM atoms(s) and thus to minimize unrealistic effects on the QM region.[374]

Most popular QM/MM boundary approaches introduce a link atom (L), typically a hydrogen atom between the QM host atom (QM_{HA}) and the MM host atom (MM_{HA}, Figure 5.2). The link atom is treated at the QM level, and sometimes with introduction of additional angular and distance constraints to keep its position fixed along the QM/MM bond. The interaction of the link atom with the surrounding MM atoms is through electrostatic terms only. In the current work, divided frontier charge (DIV) QM/MM linking scheme is employed in which the partial charge of the MM host atom is replaced by zero and distributed on to the remaining atoms of the MM host group (MM_{HG}; Figure 5.2).[374]

Other proposed boundary-atom approaches include use of pseudobond,[375] a specially parameterized connection-atom,[376] effective group potentials,[377] and quan-

tum capping potentials.[378] Alternative approaches have also been proposed such as the use of a frozen hybrid orbital[367] to terminate the QM region instead of conventional link atom. The examples of such approach include the localized self consistent field (LSCF),[379,380] frozen orbitals[381,382] and generalized hybrid orbitals.[383,384]

5.2.2 QM/MM Energy Expressions

In general, the available QM/MM approaches can be divided into two groups, an additive and a subtractive approach and depend on the way in which QM and MM energy contributions are considered.[370]

In a subtractive QM/MM approach, three calculations are carried out: (i) a MM calculation on the entire system (S); (ii) a QM calculation on the inner region (I); and (iii) a MM calculation on the inner region. The QM/MM energy of the entire system is then evaluated as the sum of (i) and (ii) after contribution from (iii) is subtracted,

$$E^{\mathrm{S}}_{\mathrm{QM/MM}} = (E^{\mathrm{S}}_{\mathrm{MM}} + E^{\mathrm{I+L}}_{\mathrm{QM}}) - E^{\mathrm{I+L}}_{\mathrm{MM}} \tag{5.1}$$

where I+L stands for the inner region saturated with link atom(s) as represented in Figure 5.2.

An example of a subtractive QM/MM approach is the IMOMM method.[385] Above represented subtractive QM/MM approach is conceptually simple since it avoids any explicit coupling terms. However, it requires a complete set of MM parameters for the inner region which may often be difficult to obtain. Further, in a subtractive QM/MM approach, the electrostatic interactions between the QM and MM regions are treated within a simple point-charge model. This approximation is serious since, (i) the charge distribution of the region I can change dynamically which can not be represented by fixed point-charge models; and (ii) the QM calculation does not incorporate the charges in the outer region which means that the effect of the polarization of the QM region due to the MM point charges is neglected.

Thus, the subtractive QM/MM approaches are not suitable for simulating chemical reactions in which the electron density is significantly influenced by electrostatic interactions with the outer region.[370]

On the other hand, in an additive approach, the QM/MM energy of the total

system, S is written as,

$$E^{\mathrm{S}}_{\mathrm{QM/MM}} = E^{\mathrm{O}}_{\mathrm{MM}} + E^{\mathrm{I+L}}_{\mathrm{QM}} + E^{\mathrm{I,O}}_{\mathrm{QM\text{-}MM}} \tag{5.2}$$

where an explicit coupling term $E^{\mathrm{I,O}}_{\mathrm{QM\text{-}MM}}$ represents the interaction terms between the inner (I) and outer (O) regions. The saturated inner region (I+L) is treated at QM level. Precise nature of the coupling terms depend on the employed approach,[370] but in general it has contributions from electrostatic (el), van der Waals (vdW), bonded interactions (b) and possibly additional constraints (cons),[386]

$$E^{\mathrm{I,O}}_{\mathrm{QM\text{-}MM}} = E^{\mathrm{el}}_{\mathrm{QM\text{-}MM}} + E^{\mathrm{vdW}}_{\mathrm{QM\text{-}MM}} + E^{\mathrm{b}}_{\mathrm{QM\text{-}MM}} + E^{\mathrm{cons}}_{\mathrm{QM\text{-}MM}} \tag{5.3}$$

The bonded terms and constraints are used to ensure proper connectivities and geometries when the QM/MM boundary lies across covalent bonds. The QM/MM van der Waals terms can be optimized to improve properties such as the distribution of MM groups around the QM region.[387] The electrostatic coupling term is of crucial importance[370] due to its significant impact on the energy (see section 5.3).

5.2.3 QM/MM Embedding Schemes

The available QM/MM approaches can be categorized into three schemes,[310,370] depending on the level of sophistication with which the electrostatic coupling between the QM charge density and the MM charge model is described.

Mechanical embedding: The QM/MM interaction is treated at the MM level and the coupling is usually limited to steric interactions. Thus the QM charge density is not influenced by the surrounding MM electrostatic environment. The mechanical embedding scheme has been used in numerous approaches including ONIOM method.[388,389] Although it is a very straightforward approach, it suffers from many limitations.[370] Most serious of which is the neglect of polarization of the QM region and hence the scheme is not suitable for reactions in which the electrostatic effect of the MM environment on the QM region is of crucial importance.

Electrostatic embedding: The limitations of mechanical embedding are eliminated in the electrostatic embedding scheme by including the polarization of the QM region by the MM charge distribution. This is accomplished by incorporating the MM point charges as one-electron terms in the QM Hamiltonian,

$$H^{\mathrm{el}}_{\mathrm{QM\text{-}MM}} = -\sum_{i}^{\text{electrons}} \sum_{M\in\mathrm{O}} \frac{q_M}{|\mathbf{r}-\mathbf{r}_M|} + \sum_{\alpha\in\mathrm{I+L}} \sum_{M\in\mathrm{O}} \frac{q_M Z_\alpha}{|\mathbf{r}_\alpha-\mathbf{r}_M|} \tag{5.4}$$

where q_M are the MM point charges, Z_α is the nuclear charge of the QM region. The index i runs over all electrons, M over all MM point charges, and α over the QM nuclei.

The MM point charges can be either those used in the force field,[368,386] or can be out of a reparameterization procedure.[390]

Polarized embedding: In polarized embedding scheme, the polarization of the MM region due to the charge distribution of the QM region is considered. The embedding scheme thus can be considered as an extension of the electrostatic embedding scheme since it rectifies the asymmetric description of the electrostatic interaction (for reviews see refs 391–393). The polarization of classical MM region can be accomplished either by employing atom-centered polarizable dipoles[367,394,395] or electronegativity equalization.[396] In the current work, a recently introduced empirical polarization model[397] is used in which the polarization within the protein environment is explicitly treated by combining a polarization-free point charge model with an interactive induced-atomic dipole model.[397]

5.2.4 SCC-DFTB/CHARMM Approach

In the current work, QM/MM calculations are carried out using SCC-DFTB[345] as the QM method with CHARMM[357,363] force field to describe the MM region. SCC-DFTB has been implemented in CHARMM for carrying out QM/MM calculations, the total energy $E^{\text{tot}}_{\text{QM/MM}}$ of the entire system is written as,[398]

$$E^{\text{tot}}_{\text{QM/MM}} = \left\langle \Psi | H_{\text{SCC-DFTB}} + H^{\text{el}}_{\text{QM/MM}} | \Psi \right\rangle + U^{\text{vdW}}_{\text{QM/MM}} + U^{\text{bonded}}_{\text{QM/MM}} + U_{\text{MM}} + U^{\text{cons}}_{\text{QM/MM}} \tag{5.5}$$

where Ψ represents the electronic wave function of the QM region. The exact form of $H^{\text{el}}_{\text{QM/MM}}$ depends on the employed QM/MM linking scheme.[374]

In the original SCC-DFTB/CHARMM implementation,[398] $H^{\text{el}}_{\text{QM/MM}}$ is evaluated as,

$$H^{\text{el}}_{\text{QM/MM}} \approx \sum_{A \in \text{MM}} \sum_{B \in \text{QM}} \frac{Q_A \Delta q_B}{|\mathbf{r}_A - \mathbf{r}_B|} \tag{5.6}$$

where the Coulomb interaction is between the MM point charge, Q_A and the Mulliken charge, Δq_B on the QM atom B. The corresponding Hamiltonian has the

form,[398]

$$\begin{aligned} H_{\mu\nu} &= H^0_{\mu\nu} + \frac{1}{2} S_{\mu\nu} \sum_{B \in \mathrm{QM}} (\gamma_{CB} + \gamma_{DB}) \Delta q_B \\ &+ \frac{1}{2} S_{\mu\nu} \sum_{A \in \mathrm{MM}} \left(\frac{Q_A \Delta q_C}{|\mathbf{r}_C - \mathbf{r}_A|} + \frac{Q_A \Delta q_D}{|\mathbf{r}_D - \mathbf{r}_A|} \right) \quad \mu \in C; \nu \in D \end{aligned} \tag{5.7}$$

where C and D denote the QM atoms on which the atomic basis functions η_μ and η_ν respectively, are located.

5.3 Implicit Solvent Models and Electrostatic Interactions

Most biochemical reactions take place in presence of the solvent environment and therefore for proper simulation of biochemical reactions, it is essential to understand the solvent effects on the chemical reaction. The exact composition of solvent environment depends largely on the particular biomolecular system of interest and can involve water molecules, ions and/or lipid membrane molecules.

In some reactions, solvent molecules are directly involved in the chemical reaction such as hydrolysis reaction. The simulation of such reactions requires an explicit representation of solvent molecules.[310,366] Most popular explicit solvent models are based on the Ewald summation with periodic boundary conditions.[399,400] However, the simulations that employ Ewald summation based methods, require treatment of large number of explicit solvent molecules for accurate description of the electrostatic interactions and thus are computationally expensive and challenging.[400] Further, the necessity of large system size in such simulations also limits the applicability of such approaches to QM/MM simulations of large biomolecules.[401–405]

Another important factor in biomolecular simulations is the accurate representation of van der Waals and electrostatic interactions. The van der Waals interactions are short range interactions and hence can be truncated beyond a certain interatomic distance. The electrostatic interactions however, are long range interactions and therefore are crucial for the structure and function of biomolecules and in particular, enzymes.[372,399,406–408]

Evaluation of long range electrostatic interactions is computationally very expensive and therefore several approaches have been proposed for faster but suffi-

ciently accurate electrostatic calculations.[407] Many approaches are based on truncation schemes, in which the electrostatic effects beyond a certain cutoff distance are completely ignored. While such approaches increase the computational efficiency, such abrupt truncation schemes lead to instabilities and artifacts in simulations.[399,409] Such artifacts can be partially rectified by employing proper shifting or switching schemes which smoothly truncate the Coulomb potential and thus improve the stability of the simulations.[399,409] Although such truncation schemes can improve the efficiency of simulations, they are inadequate for simulations in which the electrostatic interactions dominate over the short range van der Waals interactions.[399,410,411]

In many cases however, the solvent is not explicitly involved in the chemical reaction. Nevertheless, it is essential to consider its effect on the solute since it affects many properties such as equilibrium geometries and reaction rates.[412] For simulating such reactions, effects of the solvent molecules are included in an implicit manner and over the years several implicit solvent models have been proposed.[310,399,413–416]

All implicit solvent models are motivated from the Poisson equation of classical electrostatics which relates the spatial variation of the electrostatic potential ϕ to the charge density ρ as,

$$\nabla^2\phi(\mathbf{r}) = -\frac{4\pi\rho(\mathbf{r})}{\varepsilon} \tag{5.8}$$

The Poisson equation applies to a system with explicit representation of the solute with charge density ρ and implicit representation of the solvent as a homogeneous dielectric medium with dielectric constant ε. However, for systems with distance dependent dielectric constant, Poisson-Boltzmann equation must be solved,

$$\nabla\varepsilon\cdot\nabla\phi(\mathbf{r}) - \varepsilon(\mathbf{r})\lambda(\mathbf{r})\kappa^2\phi(\mathbf{r}) = -4\pi\rho(\mathbf{r}) \tag{5.9}$$

where inverse of $\kappa(\mathbf{r})$ is the space dependent ion screening length.[310,413]

In simulations with implicit solvent models, the solute (typically a protein), is represented as a macroscopic system with low dielectric constant with fixed charges and the surrounding solvent environment is represented as a dielectric continuum medium. Several models have been developed to perform the continuum electrostatics calculations based on the Poisson-Boltzmann (PB) equation.[310,399,415] Although employed in many biomolecular simulations,[399,417,418] such calculations are often very complicated and computationally expensive.

Other faster but approximate approaches such as Generalized Born (GB) models have also been developed and are used in many biomolecular simulations.[416,419] These continuum electrostatics calculations using GB models are based on the original development of Born, Onsager and Kirkwood formalism[310,415] Since these calculations are approximate, it is essential to evaluate their accuracy to describe the electrostatics of a macromolecule in solvent. This is usually done by comparison with PB calculations, although it is not clear if this is the correct approach.[399,420]

Several continuum electrostatics based simulations have indicated that the interactions of the macromolecule (or solute) with the solvent molecules in the region close to the solute (known as the first-solvation-shell) have different properties than the bulk.[310,415] The implicit solvent models that take into account the first-solvation-shell effects are widely used and are usually based on the concept of solvent-accessible surface area (SASA).[421,422]

Other approaches to model continuum electrostatics interactions employ combined implicit/explicit representation of the solvent molecules.[415,423] The charge scaling scheme[423,424] is based on such formalism in which selected charges of solvent exposed charged protein groups are scaled down following the PB calculations to mimic the bulk solvent. In the spherical solvent boundary potential (SSBP) model,[425] the effect of the first solvation shell is explicitly evaluated, while the influence of the remaining solvent environment is represented with an effective solvent boundary potential. The SSBP approach is applicable to systems with spherical macromolecules embedded in the solvent. To simulate large macromolecules of arbitrary shapes, SSBP was subsequently extended to the generalized solvent boundary potential (GSBP) approach.[426]

5.3.1 Generalized Solvent Boundary Potential

In continuum electrostatics calculations for biomolecules in a solvent environment, irregularities in the macromolecule-solvent dielectric boundaries give rise to complex reaction fields.[1] These effects can be systematically taken into account by considering the electrostatic solvation free energy contribution $\nabla W_{\text{elec}}(\mathbf{r})$ repre-

[1] In continuum electrostatics formalism, the solvent molecules are represented by a continuous electric field that represents average over all solvent degrees of freedom at thermal equilibrium. This field in the regions of space occupied by the solute is known as the reaction field.

sented as,[413]

$$\nabla W_{\text{elec}}(\mathbf{r}) = \frac{1}{2}\sum_{\alpha} q_{\alpha}\phi_{\text{rf}}(\mathbf{r}_{\alpha}) \tag{5.10}$$

where $\phi_{\text{rf}}(\mathbf{r}_{\alpha})$ is the reaction field at the position of the atomic charge q_{α}. The reaction field potential $\phi_{\text{rf}}(\mathbf{r})$ can be obtained by calculating the difference between a reference electrostatic potential computed in vacuum (ϕ_{v}), and the electrostatic potential computed in the dielectric solvent environment (ϕ_{s}).

In GSBP, the solvated biomolecule is partitioned into two regions: the inner region with a fixed dielectric constant in which a part of the biomolecule, other important molecules and a part of solvent molecules are explicitly represented, and the outer region with a variable dielectric constant that contains the remaining part of the biomolecule and solvent molecules. (Figure 5.3) The atoms in the inner region are allowed to move, while the atoms in the outer region are fixed.[426]

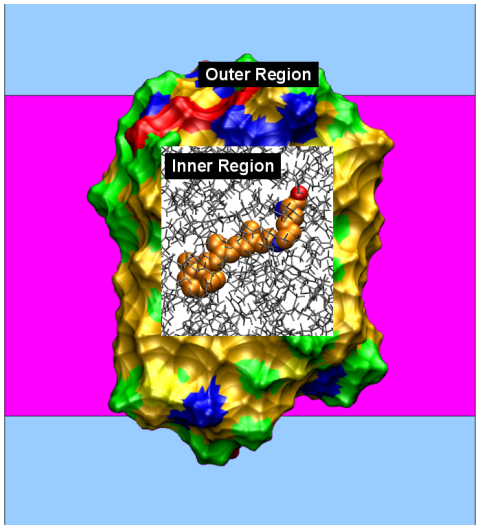


Figure 5.3: Schematic representation of the GSBP partition scheme for a biomolecule of random shape. The complete system is divided into inner and outer region. Atoms in the inner region are represented explicitly, while in the outer region consists of remaining biomolecule, the solvent molecules or membrane if applicable, with fixed positions. The effects of outer region are considered using continuum electrostatics.

Corresponding to these regions, the contributions to the electrostatic solvation free energy (Eq. 5.10) are divided into three components to include the interactions of outer-outer (oo), outer-inner (io), and inner-inner (ii) regions of the system

respectively,

$$\nabla W_{\text{elec}} = \nabla W_{\text{elec}}^{(\text{oo})} + \nabla W_{\text{elec}}^{(\text{io})} + \nabla W_{\text{elec}}^{(\text{ii})} \tag{5.11}$$

The contribution due to the outer-outer interaction ($\nabla W_{\text{elec}}^{(\text{oo})}$) arises due to the fixed charges in the outer region and is constant during the simulation. The term, $\nabla W_{\text{elec}}^{(\text{io})}$ represents the electrostatic interaction between the charges in the inner region and the reaction field due to the charges in the outer region. Hence the contribution due to the outer-inner interaction can be expressed using Eq. 5.10 as,

$$\nabla W_{\text{elec}}^{(\text{oo})} + U_{\text{elec}}^{(\text{io})} = \sum_{\alpha \in \text{inner}} q_\alpha \phi_{\text{s}}^{(\text{o})}(\mathbf{r}_\alpha) \tag{5.12}$$

where $U_{\text{elec}}^{(\text{io})}$ accounts for the direct Coulomb interaction between the inner and outer region and $\phi_{\text{s}}^{(\text{o})}(\mathbf{r}_\alpha)$ is the electrostatic potential at $\mathbf{r}_\alpha$ due to charges in the outer region only. Since the outer region atoms are fixed, their contribution is evaluated only once and saved on a set of grid points in the inner region.

The final term of Eq. 5.11 represents contribution due to the inner-inner region interaction and is expressed as,

$$\nabla W_{\text{elec}}^{(\text{ii})} = \frac{1}{2} \sum_{\alpha \in \text{inner}} q_\alpha \phi_{\text{rf}}^{(\text{i})}(\mathbf{r}_\alpha) \tag{5.13}$$

where the value of the reaction field, $\phi_{\text{rf}}^{(\text{i})}(\mathbf{r}_\alpha)$ depends on the instantaneous positions of the inner region atoms and hence requires evaluation at each step of the simulation. To avoid this computationally expensive step, the reaction field, $\phi_{\text{rf}}^{(\text{i})}(\mathbf{r}_\alpha)$ is expressed in terms of the reaction field Green's function $G_{\text{rf}}(\mathbf{r}, \mathbf{r}')$ which corresponds to the reaction field potential at $\mathbf{r}$ due to the atomic charge at $\mathbf{r}'$,

$$\phi_{\text{rf}}^{(\text{i})}(\mathbf{r}) = \int d\mathbf{r}' G_{\text{rf}}(\mathbf{r}, \mathbf{r}') \rho^{(\text{i})}(\mathbf{r}') \tag{5.14}$$

Substituting in Eq. 5.13, the inner-inner interaction contribution becomes,

$$\nabla W_{\text{elec}}^{(\text{ii})} = \frac{1}{2} \int \int d\mathbf{r} d\mathbf{r}' \rho^{(\text{i})}(\mathbf{r}) G_{\text{rf}}(\mathbf{r}, \mathbf{r}') \rho^{(\text{i})}(\mathbf{r}') \tag{5.15}$$

To achieve computational efficiency, the atomic charge distribution in the inner region is represented using a basis set expansion,

$$\rho^{(\text{i})}(\mathbf{r}) = \sum_m c_m b_m(\mathbf{r}) \tag{5.16}$$

The expansion coefficients are in turn expressed as, $c_m = \sum_n S^{-1}_{mn} Q_n$, where the overlap matrix elements S_{nm} are computed as,

$$S_{nm} = \int d\mathbf{r} b_n(\mathbf{r}) b_m(\mathbf{r}) \tag{5.17}$$

and, the elements of the vector of the generalized multipole moments Q_n are written as,

$$Q_n = \sum_{\alpha \in \text{inner}} q_\alpha b_n(\mathbf{r}_\alpha) \tag{5.18}$$

Thus, the contribution due to inner-inner region interactions becomes,

$$\begin{aligned} \nabla W^{(\text{ii})}_{\text{elec}} &= \frac{1}{2} \sum_{mn} Q_M \left[\sum_{ij} S^{-1}_{im} M_{ij} S^{-1}_{jn} \right] \\ &= \frac{1}{2} \sum_{mn} Q_M M^*_{mn} Q_n \end{aligned} \tag{5.19}$$

The elements of the generalized reaction field matrix M^* are written as,

$$M_{ij} = \int \int d\mathbf{r} d\mathbf{r}' b_i(\mathbf{r}) G_{\text{rf}}(\mathbf{r}, \mathbf{r}') b_j(\mathbf{r}') \tag{5.20}$$

The generalized reaction field matrix M^* is the numerical representation of the Green's function for the Poisson-Boltzmann equation in the b_m basis. The matrix M^* is independent of the instantaneous positions of the atoms in the inner region and hence it is calculated once and stored to increase the computational efficiency.

The basis functions are chosen depending on the shape of the biomolecule and in particular the geometry of the inner region to be simulated. For example for globular proteins, choice of a spherical inner region is advisable and correspondingly spherical harmonics are chosen, while for membrane proteins, an orthorhombic inner region may be suitable with Legendre polynomials as the choice of basis functions.[426]

5.3.2 SCC-DFTB/CHARMM Implementation of GSBP

The importance of proper treatment of long-range electrostatic interactions in biomolecular simulations has been emphasized many times in several studies.[310,372,408,415] For efficiency and accuracy of hybrid QM/MM simulations, it is essential that the long-range interactions are treated in a systematic approach.[402,427] Therefore to

improve the applicability and accuracy of SCC-DFTB/CHARMM simulations in terms of long-range electrostatics effects, the GSBP approach was implemented into the QM/MM framework few years ago.[405]

The implementation of GSBP is based on the assumption that the QM region is included in the inner GSBP region. The QM/MM contribution to the electrostatic solvation free energy can then be written as,

$$\begin{aligned} \nabla W_{\mathrm{elec}}(X_{\mathrm{QM}}, \Psi; X_{\mathrm{MM}}) &= \frac{1}{2}\sum_{mn} Q_m^{\mathrm{QM}} M_{mn} Q_n^{\mathrm{QM}} + \sum_{mn} Q_m^{\mathrm{QM}} M_{mn} Q_n^{\mathrm{MM}} \\ &+ \int d\mathbf{r}\rho^{\mathrm{QM}}(\mathbf{r})\phi_{\mathrm{rf}}^{(\mathrm{o})}(\mathbf{r}) \end{aligned} \tag{5.21}$$

where $\rho^{\mathrm{QM}}(\mathbf{r})$ is the charge density of the QM atoms and the electrostatic free energy contributes to the self-consistent solution of the QM wave function Ψ.

The first term on the r.h.s. of Eq. 5.21 represents the reaction field interaction between QM atoms, the second term describes the reaction field interaction between the QM atoms and the MM atoms of the inner region, while the last term gives the interaction between the QM atoms and the reaction field due to atoms in the outer region.

Depending on the employed scheme of the treatment of the QM/MM boundary region,[374] the electrostatic solvation free energy is modified as,

$$\begin{aligned} \nabla W_{\mathrm{elec}}(X_{\mathrm{QM}}, \Psi; X_{\mathrm{MM}}) &= \frac{1}{2}\sum_{mn} Q_m^{\mathrm{QM}} M_{mn} Q_n^{\mathrm{QM}} + \sum_{mn} Q_m^{\mathrm{QM}} M_{mn} (Q_n^{\mathrm{MM}} - Q_n^{\mathrm{EX}}) \\ &+ \int d\mathbf{r}\rho^{\mathrm{QM}}(\mathbf{r})\phi_{\mathrm{rf}}^{(\mathrm{o})}(\mathbf{r}) \end{aligned} \tag{5.22}$$

where Q_n^{EX} is defined as,

$$Q_n^{\mathrm{MM}} = \sum_{\alpha \in \mathrm{EX}} q_\alpha b_n(\mathbf{r}_\alpha) \tag{5.23}$$

Following a general QM approach, both Q_n^{QM} and the last term on the r.h.s. of Eq. 5.21 depend on the electron density of the QM region. In the case when the chosen QM method is SCC-DFTB,[345] the electron density is represented in terms of Mulliken charges[345,398] thus simplifying the GSBP calculations. The general-

ized multipole moments Q^{QM} are then evaluated as,

$$\begin{aligned} Q_m^{\mathrm{QM}} &= \int d\mathbf{r}\rho^{\mathrm{QM}}(\mathbf{r})b_m(\mathbf{r}) \\ &= \int d\mathbf{r}\Delta q^A \delta(\mathbf{r}-\mathbf{r}_A)b_m(\mathbf{r}) \\ &= \sum_{A\in \mathrm{QM}} \Delta q^A b_m(\mathbf{r}_A) \end{aligned} \tag{5.24}$$

Further using the variational terms, the Hamiltonian matrix elements $\mathrm{H}_{\mu\nu}$ of SCC-DFTB are extended to include the GSBP contributions using the following terms,

$$\begin{aligned} H_{\mu\nu}^{\mathrm{GSBP}} &= \frac{1}{2}S_{\mu\nu}\left[\Gamma_{CA}(\mathbf{r}_C,\mathbf{r}_A)+\Gamma_{DA}(\mathbf{r}_D,\mathbf{r}_A)\right]\Delta q^A \\ &+ \frac{1}{2}S_{\mu\nu}\left[\Omega(\mathbf{r}_C)+\Omega(\mathbf{r}_D)\right] \quad \mu\in C,\nu\in D \end{aligned} \tag{5.25}$$

where $S_{\mu\nu}$ is the overlap matrix element and the basis functions μ and ν are located on the QM atoms C and D respectively. The term $\Gamma_{AB}(\mathbf{r}_A,\mathbf{r}_B)$ represents the difference between the vacuum and reaction field interaction between QM atoms and is defined as,

$$\Gamma_{AB}(\mathbf{r}_A,\mathbf{r}_B)=\sum_{mn} b_m(\mathbf{r}_A)M_{mn}b_n(\mathbf{r}_B) \tag{5.26}$$

while the term $\Omega(\mathbf{r}_A)$ is written as,

$$\Omega(\mathbf{r}_A)=\sum_{mn} b_m(\mathbf{r}_A)M_{mn}(Q_n^{\mathrm{MM}}-Q_n^{\mathrm{EX}})+\phi_s^{(\mathrm{o})}(\mathbf{r}_A) \tag{5.27}$$

and it represents the direct Coulombic and reaction field contribution to the interaction between QM and the outer regions.

Chapter 6

Investigation of the Proton Release Group

Summary

This chapter summarizes the results of investigation of the PRG. The simulation setup is described in section 6.2 followed by benchmark calculations for IR bands in section 6.3. The results from the wild-type L state simulations are presented in section 6.5, and results of the mutant simulations are summarized in section 6.6. IR spectra calculations of the protonated water cluster are described in section 6.7.2 followed by conclusions of the work.

6.1 Introduction

In wild-type bacteriorhodopsin photocycle, following the first PT from the Schiff base to extracellular Asp85 during the L-to-M transition, a proton is released to the extracellular bulk with a time constant of about 80 μs.[209,428–430] Since, the newly protonated Asp85 remains protonated until the end of the photocycle,[118,238,254] the released proton must originate from a residue different than Asp85 residue.[204] Initially it was termed as XH,[204] which is more recently referred to as the proton release group (PRG). The identity of the PRG has been controversial and many residues have been proposed as the PRG.

Early FTIR spectroscopy studies of Glu204Gln and Glu204Asp mutants[215] and pK_a analysis of the primary donor Asp85[233] suggested Glu204 as the PRG. Sub-

sequent work on Glu194Asp mutants indicated that Glu204 more likely acts as an intermediate proton carrier and donates its proton to Glu194 before the proton is released to bulk.[125,216,217] Based on the observed small distance between the carboxyl groups of Glu194/Glu204 residues (3.5-3.8 Å) in the bacteriorhodopsin trimer structures,[218] the PRG was suggested to consist of Glu194 and Glu204 sharing a proton.[218] The possibility that the released proton is being shared by Glu194/Glu204 is also consistent with suggestions from FTIR data on Glu194 and Glu204 mutants.[219]

Gerwert and co-workers carried out careful analysis of the IR spectra using the wild-type bacteriorhodopsin and a collection of mutants. Based on the results of these studies, they refuted the classical proposal because the expected spectra shifts in carboxylate groups upon Glu deprotonation were not observed for the L-to-M transition in the wild-type bacteriorhodopsin. Instead, they noted the decay of a continuum band near 2000 cm^{-1} during the rise of M state, which is reminiscent of the IR signature of protonated water clusters.[431] Accordingly, it was proposed that the PRG consists of a proton is stored on a water cluster trapped in a region surrounded by Glu194, Glu204 and Arg82.[222,223]

Although it is well accepted that water in different protonation states plays a crucial role during PTs in both liquid and biomolecules,[432] assigning the excess proton to a water cluster rather than the more basic Glu residues seems counterintuitive, especially considering the close proximity of the positively charged Arg82 (Figure 2.1). Nevertheless, continuum electrostatics calculations that explicitly considered a titratable water dimer in the relevant region[221] indicated that the proton may preferentially exist in the Zundel form ($H_5O_2^+$) with the two Glu residues (194, 204) ionized to stabilize the positive charge. The continuum electrostatics calculations, however, were based on simple point charge models and therefore are not conclusive; the close distances between Glu194, Glu204 and the relevant water molecules indicate that a quantum mechanical treatment of these residues is warranted for a quantitative consideration of the protonation pattern in this region. Car-Parrinello Molecular Dynamics (CPMD) based QM/MM simulations[304] found that a protonated water cluster in the PRG region generated a continuum IR band qualitatively consistent with the experimental observation, providing further support to the idea of involving water clusters rather than amino acid side chains as the proton storage site. However, the QM region included only the proto-

nated water molecules and therefore did not allow PT to the Glu residues. In other words, these calculations[304] did not show that the protonated water cluster is stable in the bacteriorhodopsin environment and does not deprotonate to the nearby Glu residues. Further, the observation of a short distance (less than 3Å, *vide infra*) between Glu204/Glu194 side chains in high resolution X-ray structures for both the bR and the L state is consistent with the idea of trapping a proton between these two residues;[218,219] considering the uncertainty in hydrogen bonding distances even in high resolution X-ray structures,[433] however, the structural data alone do not provide a conclusive statement.

In the current work, QM/MM MD simulations were carried out with both the protonated water cluster and the Glu194/204 side chains treated with SCC-DFTB as the QM method to investigate the identity of the PRG. The results indicate that an excess proton on a water cluster is highly unstable and quickly ($\sim$1 ps) moves to become delocalized between the Glu194/Glu204 residue pair. The results from these SCC-DFTB/CHARMM simulations are also supported by expensive B3LYP/CHARMM simulations.[434]

IR spectra calculations (see section 6.2) showed that it is this delocalization that gives rise to the continuum band in the same range as observed experimentally for the wild-type bacteriorhodopsin. In addition, the vibrational frequencies for the carboxylate stretch are significantly red-shifted for Glu194/204 due to the strong interaction mediated by the delocalized proton, which explains why the characteristic spectra shifts in carboxylate groups upon Glu deprotonation were not observed experimentally in the expected region of 1700 cm^{-1} for the L to M transition in the wild-type bacteriorhodopsin.

6.2 Computational Setup

Equilibrium simulations of the PRG region with a stochastic boundary condition and explicit solvent were carried out using the CHARMM program (c32a2 version).[435] Starting from the PDB structure for the L-state (PDB code:1UCQ),[173] hydrogen atoms were added with HBUILD.[436] All basic and acidic amino acids were kept in their physiological protonation state except Asp96 and Asp115, which are known to be protonated in the L state.[259] The protein atoms were described with the all-atom CHARMM force field for proteins[437] and the water molecules were

described with the TIP3P model.[438] The 13-*cis* retinal with protonated Schiff's base was treated using the CHARMM force field with parameters taken from refs 439, 440.

In all cases the system was partitioned into a 22 Å inner region centered at N_ε atom of Arg 82 while the rest was treated as the outer region within the framework of the GSBP approach.[441] Newtonian equations-of-motion were solved for MD region (within 18 Å), and Langevin equations-of-motion were solved for the buffer region (18-22Å) with a temperature bath of 300 K.[442] All the water molecules in the inner region were subject to a weak GEO type of restraining potential to keep them inside the inner sphere with the MMFP module of CHARMM. The GEO restraining potential is in the form of a quartic polynomial on each oxygen atom in water: $k \times \Delta^2(\Delta^2 - V_P)$, with $\Delta = r - r_{off}$; k is the restraining quartic force constant (0.5 kcal/(mol·Å^4)), r is the distance of the oxygen from the center of the simulation sphere, r_{off} is the cutoff distance (22.0 - 1.5 = 20.5 Å) below which the GEO restraint is set to zero, V_P is an offset value taken to be 2.25. These parameters lead to a restraining potential on water that smoothly turns on at 20.5 Å, reaches a well at 21.5 Å with a depth of -0.625 kcal/mol, and then quickly rises to be repulsive beyond 22.0 Å. All protein atoms in the buffer region were harmonically restrained with force constants determined directly from the B-factors in the PDB file.[442] Langevin atoms were updated heuristically during the simulation to consistently treat protein groups and water molecules that may switch regions during the simulation.

Non-bonded interactions within the inner sphere were treated with an extended electrostatics model, in which groups beyond 12 Å interact as multipoles.[443] The entire system was heated gradually to 300 K and equilibrated for ∼150 ps prior to the production simulations. To account for the electrostatics between the inner and outer region atoms and the effect of solvation, the generalized solvent boundary potential (GSBP) approach developed by Im and co-workers was used .[426] The static field due to outer region atoms, ϕ_s^o, and the reaction field matrix, $\mathbf{M}$, were evaluated using Poisson-Boltzmann (PB) calculations using a focussing scheme that places a 56 Å cube of fine grid (0.4 Å) into a larger 132 Å cube of coarse grid (1.2 Å). The inner region charge density was expressed using the first 20th-order spherical harmonics with a total of 400 basis functions. The membrane environment was treated implicitly using a dielectric model with dielectric constant of 2.0 and mem-

brane thickness set to 35 Å. The optimized radii of Roux and Nina[444,445] based on experimental solvation energies of small molecules as well as the calculated interaction energy with explicit waters were adopted to define the solvent-solute dielectric boundary. Dielectric constant of 1.0 and 80.0 were used for protein and solvent respectively.

Water molecules were added to the system following the standard protocol of superimposing the system with a water sphere of 25 Å radius. Although all simulations were started with three water molecules in the PRG region based on the X-ray structure, two "bulk" water molecules were consistently observed to enter this region during the nanosecond (ns) simulations. To verify this, grand canonical Monte Carlo simulations[446] were carried out to equilibrate the water molecules while keeping the protein fixed. This procedure leads to a total of 5 water molecules in the PRG region, which is consistent with the results of MD simulations.

In most productive QM/MM simulations, the QM region included the side chains of Glu194, Glu 204 and all 3 water molecules in the PRG region; increasing the number of water molecules to five (i.e., using the GCMC result) in the QM region does not lead to significant changes in the computed IR spectra in the key (1600-2000 cm^{-1}) region.

The QM/MM boundary (between the Cα and Cβ of the Glu's) was treated using the link-atoms with the divided frontier charge scheme; previous systematic benchmark calculations[374] suggest that this scheme is satisfactory especially when the MM atom at the QM/MM frontier has very small charges, which was the case in the current study (Cα has the charge of 0.07). In several simulations (see below), a smaller QM region that includes only the PRG region water molecules and the excess proton was used (i.e., same as in the CPMD QM/MM study in ref 304) to test the proposal of having the protonated water cluster as PRG.[222,304] In those QM/MM simulations, no link atom needs to be introduced.

The QM region was treated with SCC-DFTB[345] and throughout this work, the third-order extension of SCC-DFTB[447] was used, which was found to be important for improving the proton affinity.[447] The key advantage of SCC-DFTB is its balance of computational efficiency and accuracy, which has been demonstrated in a reasonably broad set of biomolecular applications,[348,398,448,449] including bacteriorhodopsin.[119,300,450] The effectiveness of the SCC-DFTB/MM-GSBP protocol has been quantitatively evaluated using pK_a calculations with thermodynamic inte-

gration simulations.[451,452]

To construct the IR spectra during MD simulations using the SCC-DFTB/GSBP-MM potential a time step of 0.5 fs was used in order to properly sample the high frequency vibrational modes in the system. No SHAKE constraint was applied to any bond. The IR spectrum for the QM region was computed by the Fourier transform of the dipole auto-correlation function[453,454] collected from SCC-DFTB/MM MD trajectories;[455] multiple independent simulations were carried out for statistical significance. The nuclear quantum effect on the computed IR spectra was approximated with a harmonic quantum correction factor as done in the recent study of water clusters.[455]

The SCC-DFTB/MM simulations were typically carried out for $\sim$2 ns, which allow ample equilibration of the PRG region and nearby residues, especially in the context of IR spectra calculations.

6.3 Benchmark IR bands of Gas-phase Models

6.3.1 Aspartic and Glutamic Acid

The IR bands for the aspartic acid (Ac-Asp-OCH_3) and glutamic acid (Ac-Glu-OCH_3) were calculated by treating all the atoms with the QM method (see Tables 6.1 and 6.3). To investigate the effect of QM/MM partition on the IR bands, a part of the molecule (side chain atoms) was treated with QM as in the QM/MM scheme (see Table 6.2). The QM/MM link atom was introduced between C_β and C_α atoms. The IR bands in the DFTB/CHARMM framework were calculated using VIBRAN module with different link atom schemes (EXGR and DIV) and using default 2nd and 3rd order of DFTB method.

Table 6.1: Aspartic acid[a]

Mode	Vibran (2nd order)	Vibran (3rd order)	DFTB (2nd order)	DFTB (3rd order)	PBEPBE 6-311g(2d,2p)	PBEPBE 6-311++g(2d,2p)
ν_{COOH}	1769.41	1743.87	1769.74	1762.44	1778.91	1759.25
ν_{OH}	3663.50	3666.31	3666.20	3672.55	3636.44	3641.11

[a] Acetyl and OCH_3 used as protecting groups.

Table 6.2: Aspartic acid with QM/MM link atom between CB and CA using EXGR and DIV linking schemes.

Mode	Vibran (2nd order dftb) EXGR[a]	Vibran (2nd order dftb) DIV[a]	Vibran (3rd order dftb) EXGR	Vibran (3rd order dftb) DIV
ν_{COOH}	1777.43	1781.35	1758.30	1766.50
ν_{OH}	3666.19	3665.60	3673.79	3671.94

[a] for description of these link atom schemes see ref 374

Table 6.3: Glutamic acid

Mode	Vibran (full-dftb)	Vibran (full-dftb) (3rd order)	Vibran qm/mm redufix [a]	DFTB (stand alone) (Henrik))	PBEPBE 6-311g (2d,2p)	PBEPBE 6-311++g (2d,2p)
ν_{COOH}	1762.99	1744.70	1777.15	1769.05	1767.85	1747.29
ν_{OH}	3656.47	3666.20	3663.96	3664.45	3632.44	3638.22

[a] DFTB (2nd order)

Table 6.4: Glutamic acid with QM/MM link atom between CB and CA

Mode	Vibran (2nd order dftb) EXGR	Vibran (2nd order dftb) DIV	Vibran (3rd order dftb) EXGR	Vibran (3rd order dftb) DIV
ν_{COOH}	1764.47	1765.18	1752.47	1753.52
ν_{OH}	3659.83	3659.45	3666.44	3665.78

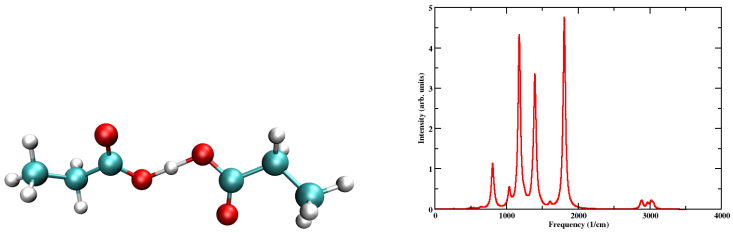
As has been shown in the earlier investigation of various link atom schemes,[374] the error in the deprotonation energies for these amino acids with the DIV linking scheme was minimum when the link atom was in between C_β and C_α. Interestingly, with the same linking scheme and 3rd order DFTB as QM method the

ν_{COOH} band red-shifted from aspartic acid (1766.5 cm^{-1}) to glutamic acid (1753.5 cm^{-1}) to by about 13 cm^{-1}. Such a red-shift has also been observed in case of bacteriorhodopsin when the Asp85 and Asp96 residues were mutated with Glu residue.[161,237,238]

Thus, although absolute IR bands were red-shifted with DFTB (3rd order), the order of red-shift between aspartic acid to glutamic acid calculated with DFTB (3rd order) is consistent with the experimental observation.

6.3.2 Gas Phase Models of Propionic Acids

It has been shown that in comparison with acetic acid, the propionic acid represents a better model for investigating the IR spectra of amino acids such as Asp or Glu.[456] Therefore the IR bands of propionic acids in various scenarios were investigated (see Tables 6.5, 6.6, 6.7, and 6.8, and Figures 6.1, 6.2, 6.3, and 6.4). The gas-phase IR bands were calculated using DFTB (2rd order), DFTB (3rd order) and DFT using PBE functional.



(a) (b)

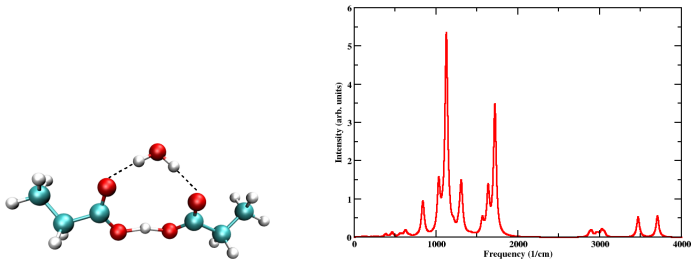
Figure 6.1: (a) Propionic acids with shared proton, (b) IR spectra

Table 6.5: Propionic acids with shared proton (Figure 6.1a)

Mode	DFTB (2nd order)	DFTB (3rd order)	PBEPBE 6-311g(2d,2p)	PBEPBE 6-311++g(2d,2p)
ν_{COOH}	1603.38	1587.18	1681.94	1648.02
ν_{OH}	1802.10	1789.03	1730.69	1702.22

In the first scenario where propionic acid share a proton (see Figure 6.1a), the

ν_{COOH} band was red-shifted from 1769.2 cm^{-1} to 1603 cm^{-1}. Also the ν_{OH} band was red-shifted from 3659.3 cm^{-1} to 1802 cm^{-1} (see Table 6.5). The corresponding IR spectra calculated with DFTB is shown in Figure 6.1b. The strong sharing of proton between the acids thus significantly red-shifted both ν_{COOH} and ν_{OH} bands. The situation was more amplified with addition of a water as in the next case (see Figure 6.2a). The new water molecule formed hydrogen bonds with the oxygen atoms of the double bond and thus reduced the double-bond character of the acid and red-shifted the IR bands. The situation is more like in bacteriorhodopsin where the Glu194/Glu204 residue pair share hydrogen bonds with water molecules. Important point to note here is that when the proton was shared between the acids, the ν_{OH} red-shifted significantly to approximately 2000 cm^{-1} and gave rise to a broad peak in the 1800-2000 cm^{-1} region as can be seen in Figure 6.2b.



(a) (b)

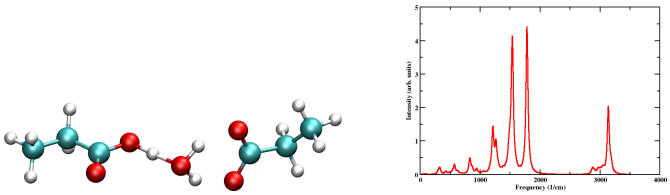
Figure 6.2: (a) Propionic acids with shared proton plus extra water, (b) IR spectra

Table 6.6: Propionic acids with shared proton and extra water (Figure 6.2a)

Mode	DFTB (2nd order)	DFTB (3rd order)	PBEPBE 6-311g(2d,2p)	PBEPBE 6-311++g(2d,2p)
ν_{COOH}	1597.40	1578.62	1663.83	1644.53
ν_{OH}	1780.91	1760.53	1758.02	1677.77

When a water was shared between the two propionic acids (see Figure 6.3a), two sharp peaks were seen corresponding to ν_{COOH} and ν_{OH} bands and were strongly coupled to the other water bands. In presence of additional water however the situation is very interesting. Again in this case, two sharp peaks can be seen cor-

responding to the ν_{COOH} (1661.4 cm^{-1}) and ν_{OH} (2027.38 cm^{-1}) bands. These bands were also strongly coupled to other water bands. Thus presence of a shared water between the acids does not give rise to any continuum band (see Figures 6.3b and 6.4b).

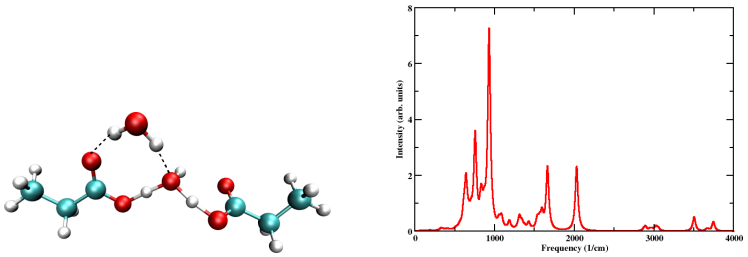


(a) (b)

Figure 6.3: (a) Propionic acids with shared water, (b) IR spectra

Table 6.7: Propionic acids with shared water (Figure 6.3a)

Mode	DFTB (2nd order)	DFTB (3rd order)	PBEPBE 6-311g(2d,2p)	PBEPBE 6-311++g(2d,2p)
ν_{COOH}	1600.20	1585.51	1653.86	1641.52
ν_{OH}	1837.79	1810.26	1939.77	1950.86



(a) (b)

Figure 6.4: (a) Propionic acids with shared water and extra water, (b) IR spectra

Table 6.8: Propionic acids with shared water and extra water (Figure 6.4a)

Mode	DFTB (2nd order)	DFTB (3rd order)	PBEPBE 6-311g(2d,2p)	PBEPBE 6-311++g(2d,2p)
ν_{COOH}	1656.09	1648.26	1682.04	1689.22
ν_{OH}	2041.34	2024.25	2023.67	1932.24

6.4 Overview of Wild-type Crystal Structures

An overview of distances between key protein groups in the PRG region in available crystal structures is shown in Table 6.9. The position of key residues (Arg82, Glu194/204), the hydrogen bonding pattern between them and the number of resolved water molecules are largely consistent among available L state structures (Figure 2.3).

Table 6.9: Comparison of various bR and L-state crystal structures of the region around Glu194/Glu204 residues in bacteriorhodopsin.[a]

State	PDB code (Res.[b])	Glu194(O^{ϵ})-Glu204(O^{ϵ}) distance (Å)	No. of waters around Glu194-Glu204 residues	Ser193(O^{γ}) -Glu204(O^{ϵ}) distance (Å)
bR	1C3W (1.55 Å)[21]	3.0	3	2.6
bR	1QHJ (1.90 Å)[76]	2.4	4	3.4
bR	1KGB (1.65 Å)[207]	2.5	3	2.6
bR	1IW6 (2.30 Å)[139]	2.5	3	2.9
bR	1C8R (1.80 Å)[210]	2.6	3	2.6
L	1E0P (2.10 Å)[171]	3.1	3	4.2
L	1O0A (1.62 Å)[172]	3.1	3	2.6
L	1UCQ (2.40 Å)[173]	2.6	3	3.0
L	1VJM (2.30 Å)[135]	2.4	3	3.4
L	2NTW (1.53 Å)[174]	4.2	3	2.3

[a] The shortest distances involving the O^{ϵ} atoms in Glu194:Glu204 are reported. [b] Resolution of the X-ray structure in Å.

In particular, as shown in Table 6.9, the pair of conserved Glu residues (Glu 194/204) in the PRG region are spatially very close in most X-ray structures, with the shortest O-O distances substantially below 3 Å. Among the bR resting state structures, only that proposed by Luecke and co-workers[21] has the distance slightly longer 3 Å.

Among the L-state structures, in the recently proposed structure of Lanyi and co-workers,[174] the significantly longer Glu194:Glu204 distance is due to the extracellular movement of the Arg82 side chain. The authors have stated that some of the changes observed in L structure resemble those of a structure after proton release (i.e., the M state) although small in magnitude,[174] and can be a reason for the observed longer Glu194:Glu204 distance.

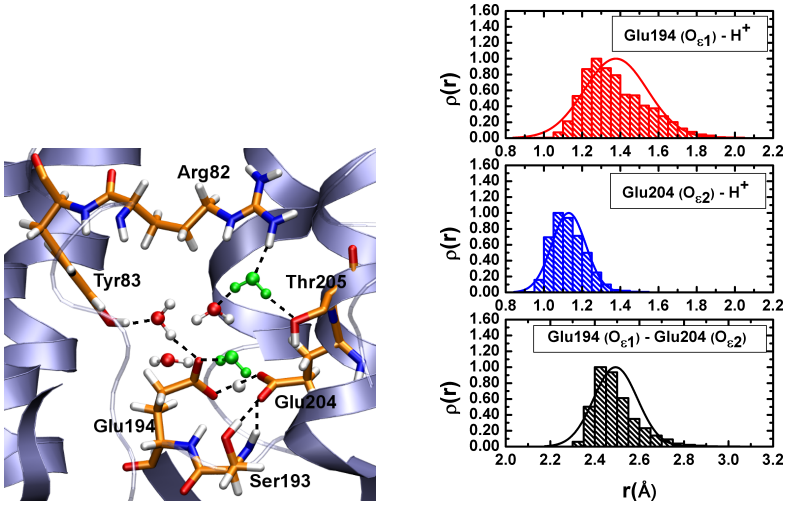
The high resolution structure 1O0A[172] has the retinal in the 13-*cis*, 15-*syn* conformation; previous study[300] of PT between the Schiff base and Asp 85 found that such a configuration is not suited for active proton pumping.

The extracellular orientation of Arg82 side chain is observed in the L state structure proposed by Royant and co-workers,[171] does not agree with other L state structures (Figure 2.3). Assignment of this structure to the L intermediate has been controversial due to contamination with the K and M intermediate states[132, 176, 177] and therefore the extracellular orientation of Arg82 as well as configuration of the two gluatamate residues Glu194 and Glu204 are probably more M-like rather than bR- or L-like as in Figure 2.3 (structure in orange color).

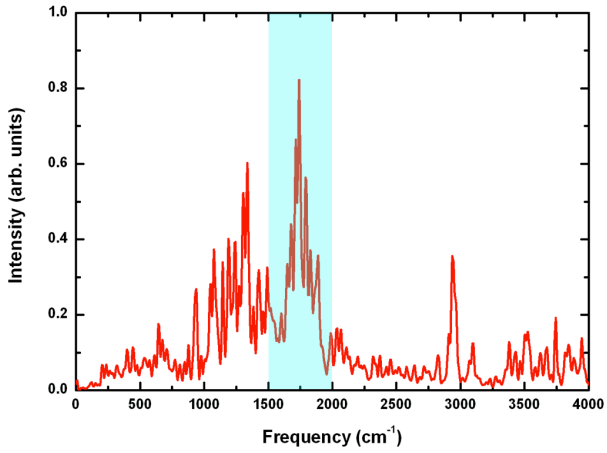
Consequently, the L structure proposed by Kouyama and co-workers[173] was chosen as the starting structure of the L-state simulations in this work. Although the resolution of this structure is not very high (2.4 Å), the fact that all key residues and water molecules in the PRG region are in the proper locations suggests that this is a sensible choice for the current purpose.

6.5 Structural and Spectral Features of the Wild-type L state

During the SCC-DFTB/CHARMM simulations, the excess proton on the water cluster in the PRG region is highly unstable and quickly moves ($\sim$ 1ps) to become delocalized between the Glu194/Glu204 residue pair (Figure 6.5a) and remains there for the rest of the simulation.



(a) (b)



(c)

Figure 6.5: (a) A snapshot of the PRG region indicate that the excess proton is shared between Glu194 and Glu204. (b) Histograms for key distances involving the excess proton and Glu194/Glu204 side chains. (c) Computed IR spectra for the QM atoms (Glu194, Glu204 side chains, excess proton, and three active site water molecules) indicate the continuum band in the 1800-2000 cm^{-1} region (highlighted in blue).

The simulations showed a strongly shared proton between the Glu194 and Glu204 residues with average distance between these residues around 2.4 Å-2.5 Å (Figure 6.5b) which is in good agreement with distances found in the crystal

structures (Table 6.9). The histograms for the minimal distance between the excess proton and Glu194/Glu204 residues (Figure 6.5b) indicate that the excess proton is shared with a slight preference over Glu204.

The Glu194 residue exhibits slightly more flexibility and forms hydrogen bonds with preferentially the waters in the active site. Both side chain oxygen group and the backbone nitrogen group of the Ser193 residue form hydrogen bonds with Glu204 oxygen atom that does not share the proton (Figure 6.5a) and stabilize the partial negative charge. Tyr83 and the active site water molecules form hydrogen bonds with the other oxygen atom of the Glu194 residue.

Although all the simulations were started with three active site water molecules, during the simulations two additional water molecules (shown in green in Figure 6.5a) were consistently observed to move in the PRG region during the ns scale simulation.

The computed IR spectra of the QM atoms at the SCC-DFTB/CHARMM level clearly show the continuum band in the 1800-2000 cm^{-1} region (Figure 6.5c) and agree well with the observed FTIR spectra.[220] To confirm that the continuum band in the 1800-200 cm^{-1} region arises due to the delocalized nature of the excess proton, additional simulations were carried out in which only one of Glu194 or Glu204 was treated with QM and the remaining Glu with MM and the results are as follows:

(i) QM/MM simulations (2ns trajectories) were carried out starting from L state structure in which Glu204 was treated with MM and the QM region consisted of Glu194 side chain with the excess proton localized on Glu194, and the active site water molecules. Throughout the dynamics, the excess proton remained localized on Glu194, while the negative charge on Glu204 is stabilized by active site water molecules and Ser193 (Figure 6.6a). Although the simulations were initiated with three water molecules, two additional bulk waters were consistently observed in the PRG region (represented in green in Figure 6.6a).

In all the simulations, Glu194:Glu204 distance remained $\sim$2.6 Å (black panel in Figure 6.6) which is consistent with Table 6.9. The corresponding IR spectra (Figure 6.6c) indicate absence of the continuum band in the 1800-2000 cm^{-1} region. Instead a sharp peak at 1657 cm^{-1} is observed which corresponds to the C=O band (ν_{COOH}) of protonated Glu194.

(ii) Another set of QM/MM simulations (2ns trajectories) were carried out in

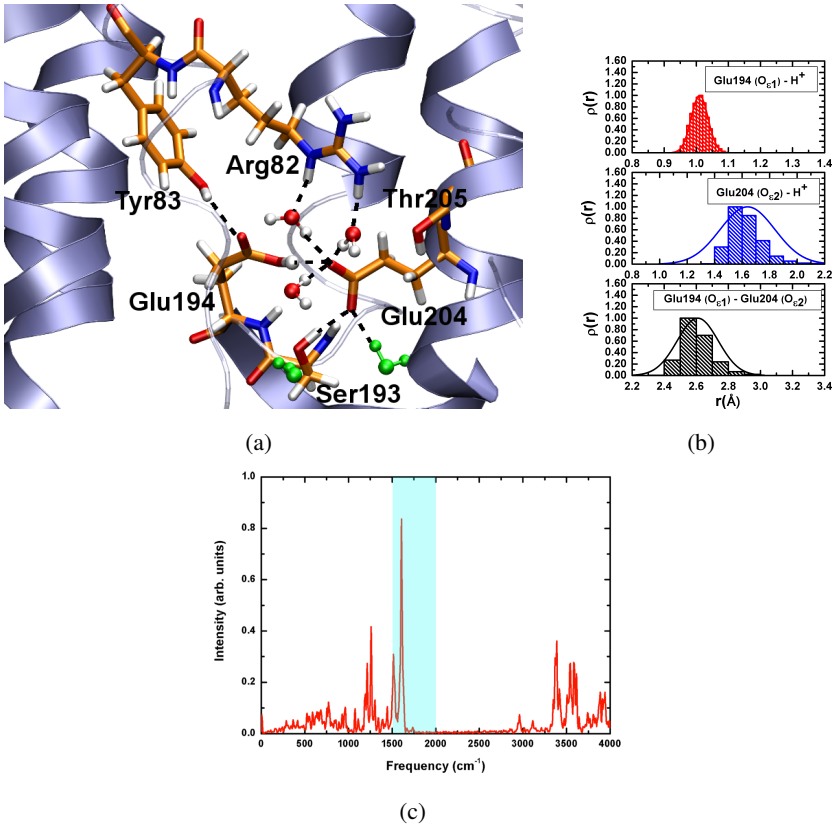


Figure 6.6: Representative results from QM/MM simulations for wild-type L state in which Glu204 was treated with MM and Glu194 side chain with excess proton and three active site water molecules were treated with QM. (a) A snapshot of the PRG region. The negative charge of Glu204 is stabilized by active site water molecules and Ser193. The bulks water molecules are shown in green. (b) The distance histogram indicate that the excess proton is localized on Glu194 (panel in red) while the Glu194:Glu204 distance is maintained around 2.6 Å. (c) Corresponding IR spectra shows no continuum band in the region 1800-2000 cm^{-1}.

which Glu194 was treated with MM instead of Glu204. The QM region consisted of Glu204 side chain with excess proton and the three active site water molecules. As in case (i), the continuum band in the 1800-2000 cm^{-1} region is absent in the IR spectra (Figure 6.7c), and the excess proton remained localized on Glu204. The negative charge on Glu194 is stabilized by active site water molecules (Figure 6.7a).

The results from cases (i) and (ii), together with the results of the simulations

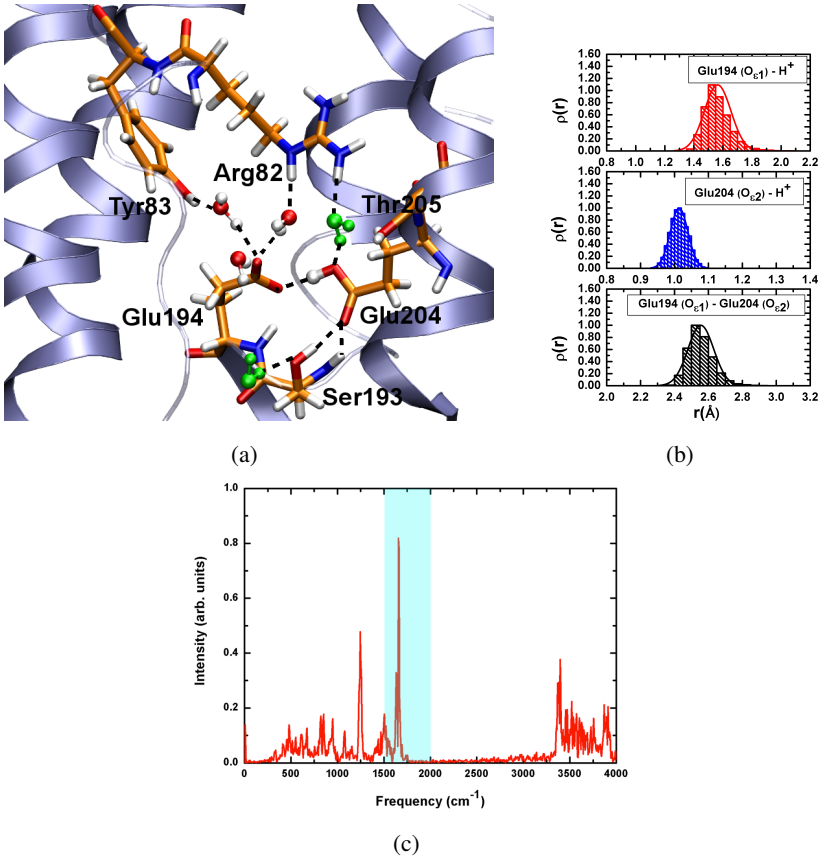


Figure 6.7: Representative results from QM/MM simulations for wild-type L state in which Glu194 was treated with MM and Glu204 side chain with excess proton and three active site water molecules were treated with QM. (a) A snapshot of the PRG region. The negative charge of Glu194 is stabilized by active site water molecules. The bulks water molecules are shown in green. (b) The distance histogram indicate that the excess proton is localized on Glu204 (panel in blue) while the Glu194:Glu204 distance is maintained $\sim$2.6 Å (panel in black). (c) Corresponding IR spectra shows no continuum band in the region 1800-2000 cm^{-1}.

with large QM region (see section 6.5) indicate that the continuum band in 1800-2000 cm^{-1} originates due to the shared nature of the excess proton and it is absent when one of the Glu194/204 residues is treated with MM.

Another important consequence of the strong hydrogen bonding interaction is that the ν_{COOH} band of Glu194/Glu204 is significantly red-shifted to below 1700 cm^{-1} which coincides with amide bands and therefore are difficult to identify un-

equivocally. Such strong red-shifts of IR bands are in agreement with gas-phase models with a shared proton at both the SCC-DFTB and DFT level (see section 6.3). This can explain why the ν_{COOH} bands have not been identified in the previous FTIR studies.[219,220]

6.6 Structural and Spectral Features of Mutant L Structures

In order to understand the role protein groups in the PRG region on the continuum band, spectral features of several mutant structures have been studied.[215,220,222] Consequently, QM/MM simulations for several mutant structures were carried out in the current work to test the robustness of the proposed model. The mutant structures were generated after *in-silico* mutations of particular protein group under investigation.

6.6.1 Glu194Asp Mutant Spectra

The simulations for wild-type L structure indicate a strong hydrogen bond between Glu194 and Glu204 ($\sim$2.5 Å; section 6.5). In case of Glu194Asp mutant simulations, the distance between Asp194 and Glu204 is $\sim$ 2.7 Å indicating a somewhat weaker hydrogen bond as a result of the mutation. During the simulations the excess proton remained localized on Glu204 (Figure 6.8b). The negative charge of Asp194 is stabilized by Tyr83, Ser193 and surrounding water molecules. The presence of strong hydrogen bonds between Asp194 with the surrounding groups reduces its flexibility as a result of which the excess proton remained localized on Glu204 (Figure 6.8a).

The experimental IR studies[222] which suggested that in the resting state of Glu204Asp mutant, Glu204 is protonated with a characteristic ν_{COOH} band at 1712 cm^{-1}. Hence, to compare this assignment, the IR spectra was computed using only the dipole moment of COOH group of Glu204 in Glu194Asp mutant and it shows a high peak at 1645 cm^{-1} (Figure 6.8d). The results from QM/MM simulations are thus qualitatively in agreement with the experimental results.

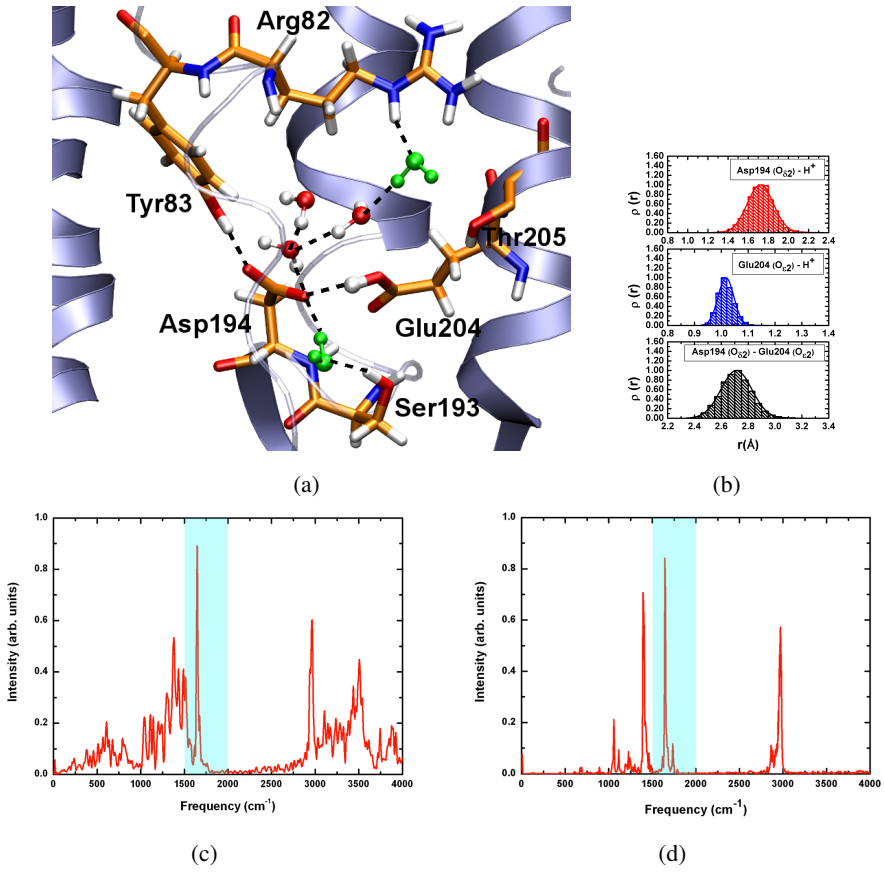


Figure 6.8: Representative results from QM/MM simulations for Glu194Asp mutant L structure. (a) A snapshot of the PRG region indicate that the excess proton is localized on Glu204. (b) The distance histograms indicate that the excess proton is largely localized on Glu204 (blue panel). (c) Computed IR spectra indicates that the continuum band in the 1800-2000 cm^{-1} region is absent. (d) Calculated IR spectra of COOH for Glu204 shows a distinct sharp peak at 1645 cm^{-1}.

6.6.2 Glu204Asp Mutant Spectra

The simulations of Glu204Asp mutant structure indicate that although the excess proton remained localized mostly on Asp204 (Figure 6.9a), higher flexibility of Glu194 results in weak sharing of proton with Asp204 as can be seen from the distance histograms (Figure 6.9b) which gives rise to a weak continuum band in the 1800-2000 cm^{-1} region (Figure 6.9c). The IR spectra for the Asp204 carboxylate group shows a high peak at 1686 cm^{-1} (Figure 6.9d) similar to Glu194Asp mutant

simulations.

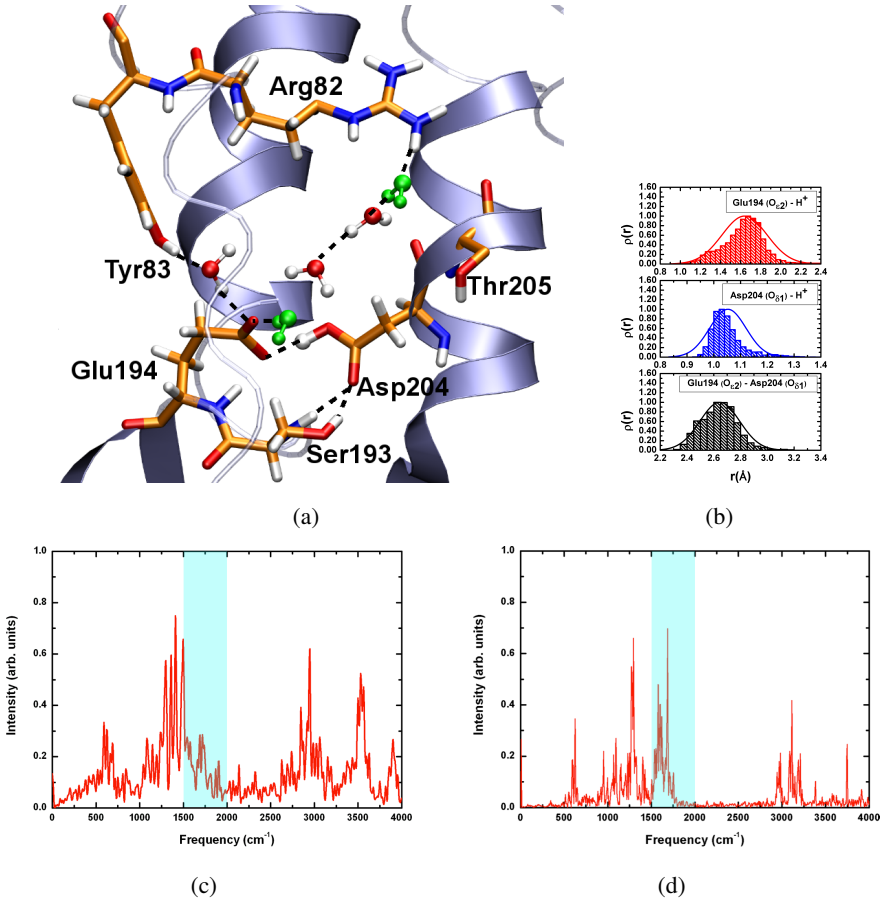
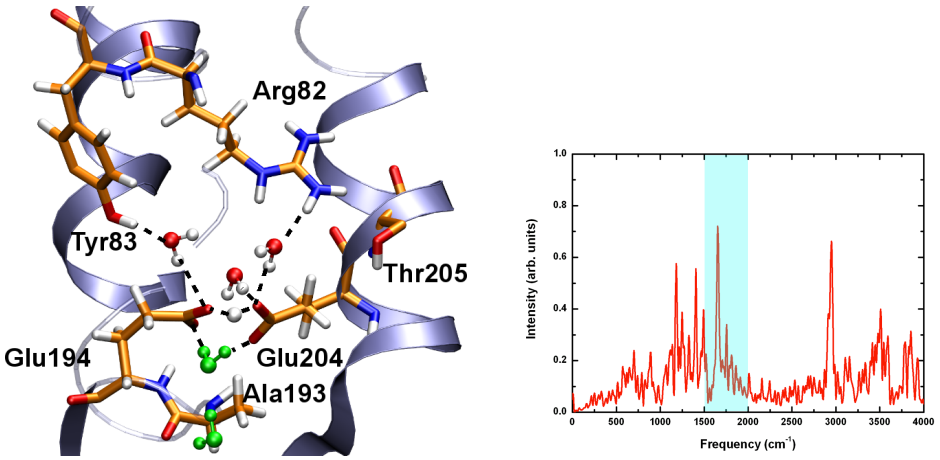


Figure 6.9: Representative results from QM/MM simulations for Glu204Asp mutant L structure. (a) A snapshot of the PRG region indicate that the excess proton is localized on Asp204. (b) The distance histograms indicate that the excess proton is weakly shared between Glu194 and Asp204. (c) Computed IR spectra indicates that a weak continuum band in the 1800-2000 cm^{-1} region is present. (d) Calculated IR spectra of COOH for Asp204 shows a high peak at 1686 cm^{-1}.

Unlike Glu194Asp mutant simulations in which the proton remained localized on Glu204, in case of Glu204Asp simulations, the proton occasionally jumps to Glu194 but returns back to Asp204. The negative charge on Glu194 is stabilized by Tyr83 and active site water molecules. The hydrogen bond between Ser193 and Asp204 is stable during most part of the trajectory.

6.6.3 Ser193Ala Mutant Spectra

The continuum band observed in the wild-type simulations is also observed in the simulations with Ser193Ala mutant (see Figure 6.10) as a result of the shared proton between the two glutamate residues. The distance between the proton sharing oxygen atoms of Glu194 and Glu204 residues in these simulations varies around 2.7 Å-2.8 Å as contrast to 2.4 Å-2.6 Å of the wild-type simulations and thus results in a weaker continuum band.



(a) (b)

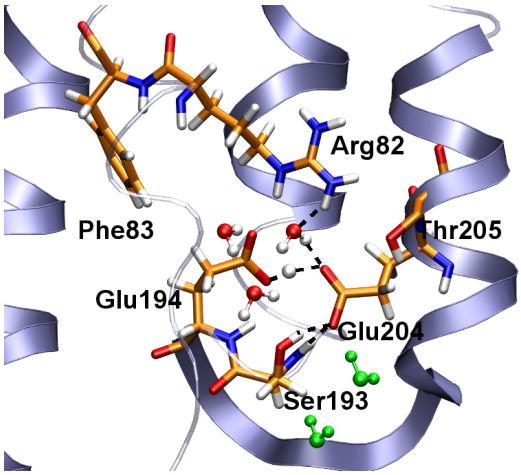
Figure 6.10: (a) Active site geometry from Ser193Ala mutant simulations; (b) Computed IR spectra for Ser193Ala mutant indicates that a continuum band in the 1800-2000 cm^{-1} region.

This is due to absence of a strong hydrogen bond between Ser193 side chain oxygen atom with the Glu204 residue as a result of mutation at position 193 to alanine residue that replaces the hydroxyl group of the serine with the methyl group. As a consequence of the absent hydrogen bond, Glu204 residue becomes more flexible as compared with to its wild-type conformation and hence thus affects the continuum band.

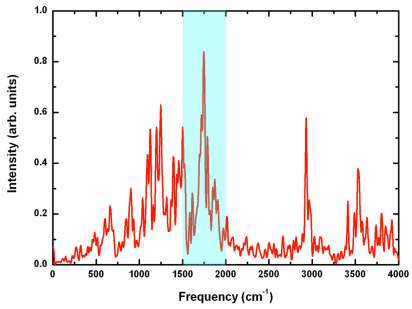
In spite of the absence of the hydroxyl group, active site water molecules form hydrogen bond with Glu204 and weakly stabilize the Glu204 conformation such that the proton remains shared.

6.6.4 Tyr83Phe Mutant Spectra

In the wild-type simulations, Tyr83 residue forms hydrogen bond with Glu194 residue either directly or via intermediate water molecules. Experimental studies[222] have indicated that replacement of Tyrosine at position 83 by Phenylalanine does not significantly affect the continuum band. The simulations with Tyr83Phe mutant carried out in this work are thus in agreement with this observation (see Figure 6.11).



(a)



(b)

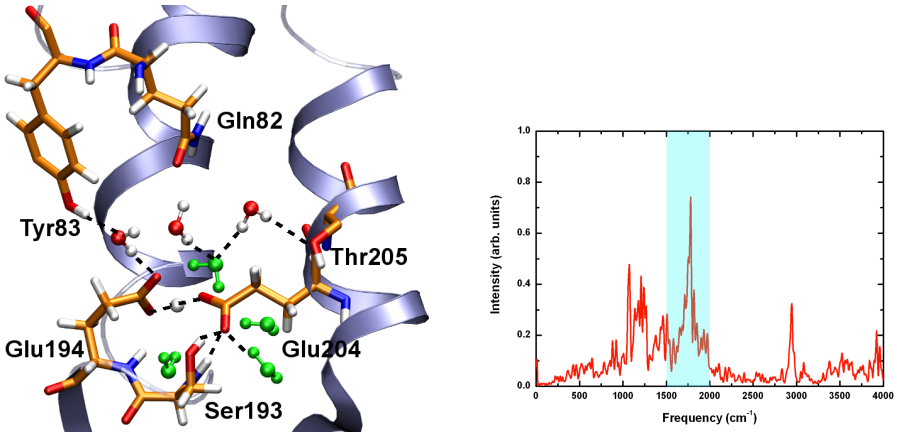
Figure 6.11: (a) Active site geometry from Tyr83Phe mutant simulations; (b) Computed IR spectra for Tyr83Phe mutant indicates a continuum band in the 1800-2000 cm^{-1} region.

In the simulations, the proton remained delocalized between the Glu194/Glu204 residues with average distance between the proton sharing oxygen atoms around

2.4 Å-2.7 Å. Thus, the absence of one hydrogen bond of Glu194 to Tyr83 does not significantly affect the continuum band. The absence of hydrogen bond of Tyr83 with Glu194 is partially compensated by the active site water molecules while Ser193 forms stable hydrogen bond with Glu204 residue.

6.6.5 Arg82Gln Mutant Spectra

The simulations with Arg82Gln mutant are most interesting. The experimental studies[222] have indicated that in such a mutant, the early proton release to the extracellular bulk is blocked and the proton release to the bulk occurs after the proton recovery from the cytoplasmic side. Further, no continuum band was observed in such mutant. However, it is not clear from the experiments, whether the absence of the continuum band is due to the localization of the excess proton on glutamates or due to the unprotonated PRG.



(a) (b)

Figure 6.12: (a) Active site geometry from Arg82Gln mutant simulations; (b) Computed IR spectra for Arg82Gln mutant indicates that a continuum band in the 1800-2000 cm^{-1} region.

The simulations with Arg82Gln mutant indicate that the excess proton remains delocalized between the two glutamates and gives rise to a continuum band in the 1800-2000 cm^{-1} region (Figure 6.12). The stability of the two negatively charged Glu residues is due to the delocalized proton and hydrogen bonding interactions between the two Glu residues and surrounding protein groups Tyr83, Thr205, Ser193 along with active site water molecules.

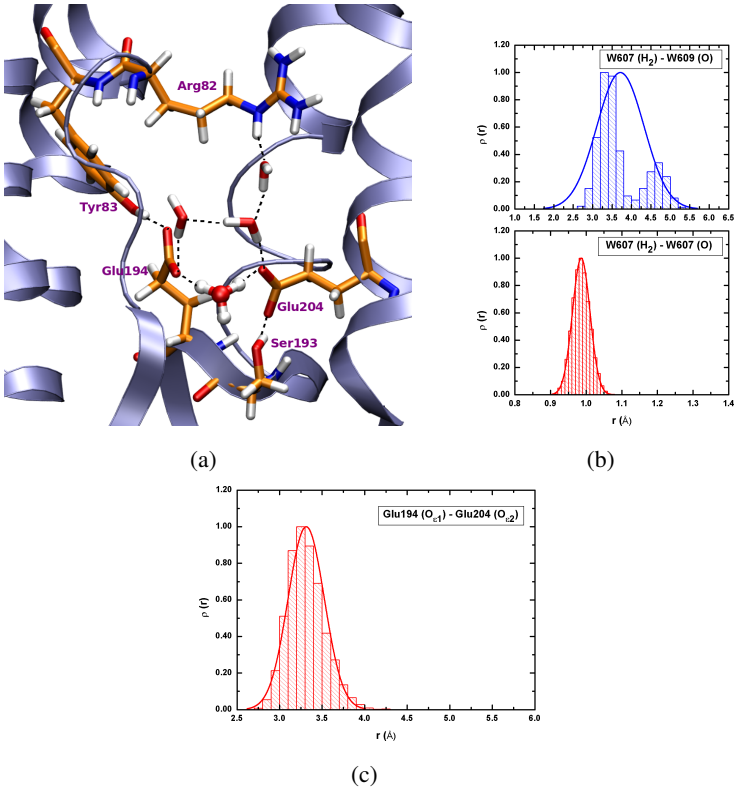
During the simulations atleast four bulk water molecules were persistently observed in the active site (Figure 6.12a) which is in contrast to the wild-type and mutant simulations described earlier. This may be due to availability of extra space as a result of mutation of bulky arginine residue to less bulky glutamine at position 82.

6.7 Simulations with a Protonated Water Cluster as PRG

To independently evaluate the proposal that the PRG is a water cluster near the conserved Glu194/204 residues,[222,223,304] SCC-DFTB/MM simulations were carried out using a small QM region that includes only the excess proton and the water molecules in the active site. To be able to compare directly with the QM/MM CPMD study,[304] four water molecules were included in the PRG region and treated as QM. Based on discussions presented earlier, the number of included water molecules is between the number of water molecules observed in various X-ray structures (section 6.4) and the GCMC results.[434] Similar to the large QM simulations discussed earlier, each set of small QM simulations was done with five independent trajectories; each trajectory lasts 1 ns, somewhat shorter than the large QM simulations.

6.7.1 Without Constraints on the QM Water Molecules

Without any specific constraint on the QM water molecules, the excess proton is mainly observed in an Eigen-like form (see a snapshot in Figure 6.13a and the distance histograms in Figure 6.13b), stabilized by the two negatively charged Glu residues. As a result, the distance between the anionic Glu residues has a distribution centered around $\sim$3.3 Å (Figure 6.13c) which is substantially longer than the distance of $\sim$2.5-3.0 Å observed in various crystal structures (see section 6.4). Considering the typical uncertainty in hydrogen bonding distances in X-ray structures, which is about 10% of the respective resolution,[433] the X-ray structures are more consistent with the active site geometry observed in the large QM region simulations.



(a) (b)

(c)

Figure 6.13: Representative results for the SCC-DFTB/MM simulations (five independent 1ns trajectories) with a small QM region that includes only water molecules in the PRG region. (a) A snapshot that illustrated the Eigen-like ion in the PRG region; (b) Distribution for distances involving the excess proton and neighboring oxygen atoms; (c) Distance distribution of the closest Glu194 - Glu204 oxygens.

6.7.2 With Constraints on the QM Water Molecules

Since the protonated water cluster was reported to exist mainly in the Zundel form in the QM/MM CPMD study,[304] additional simulations were carried out in which weak NOE constraints were applied to the QM water molecules to favor the Zundel form configuration (see Figure 6.14a for a snapshot and Figure 6.14b for distance histograms that support the Zundel population). The presence of the Zundel ion increases the distance between the Glu194/204 residues from ∼2.5 Å as in the crystal structure to ∼4.4 Å (see Figure 6.14c for the distribution). During the simulations, two additional bulk water molecules (shown in green in Figure 6.14a)

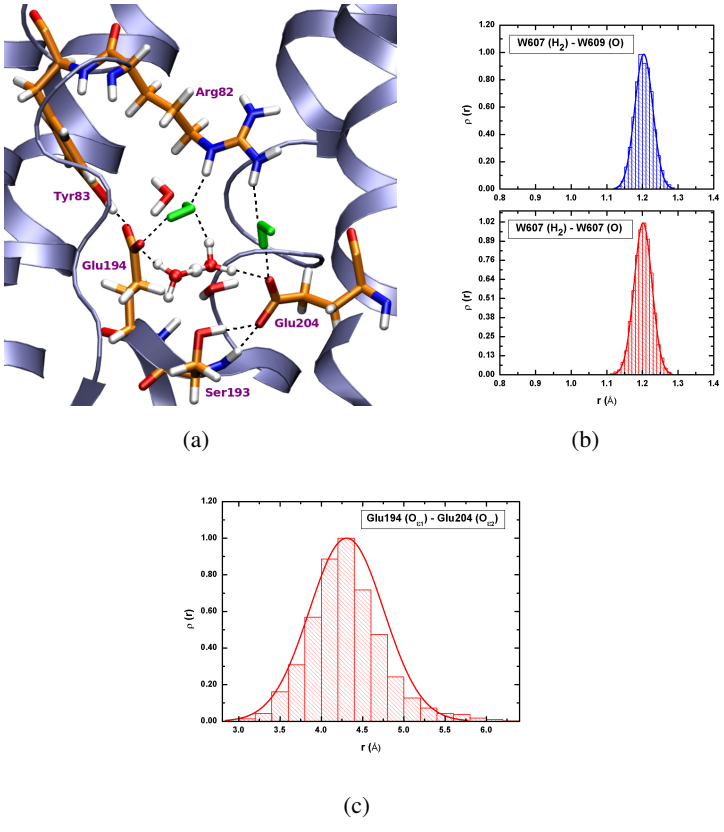


Figure 6.14: Same as Figure 6.13, but with the additional NOE constraints between two QM water oxygen atoms to favor the Zundel form of the protonated water cluster.

become involved in the hydrogen bonded network to further stabilize the anionic Glu194/204 residues.

In summary, at the SCC-DFTB/MM level, regardless of the configuration of the protonated water cluster (Eigen-like vs. Zundel), the active geometrymetry especially with respect to the conserved Glu194:Glu204 distance is largely different that those observed in various X-ray structures.

6.8 Discussion

In summary, QM/MM simulations carried out in this work indicate that a model in which the excess proton is stored in a delocalized fashion between the two con-

served Glu residues produces IR spectral features consistent with experiments for both the wild-type and several key mutants of bacteriorhodopsin. This model is chemically more intuitive compared to the proposal[222] that involves a protonated water cluster near a positively charged Arg and two more basic Glu residues; the water cluster model also leads to structural features less consistent with available X-ray data.

FTIR studies[222] observed a very weak change in the continuum band during L→M transition in Arg82Gln mutant. The reason for this change was argued to be due to the merging of water clusters.[222,457] QM/MM simulations of Arg82Gln in the current work indicate the mutation of arginine to glutamine at position 82 has little effect of the delocalization of the excess proton between Glu194/Glu204 residues and consequently on the continuum band. However, it should be noted that the photocycle of Arg82Gln mutant is considerably different from the wild-type photocycle.[214] Further changes in the hydrogen bonding pattern of Arg82 has been observed in previous studies.[224] Therefore, further investigations will be carried out in future in which Arg82 will be included in the QM region to understand the changes in the hydrogen bonding pattern of Arg82 and its effect on the continuum band.

Conceptually, the model presented here does not disqualify the functional importance of water molecules in PT. Rather, the finding emphasizes that proton delocalization is not limited to cases involving water molecules as in the commonly discussed Zundel/Eigen cations. In fact, using an "intermolecular proton bond",[458] or a "low-barrier hydrogen bond",[459,460] to stabilize two otherwise repulsive groups (e.g., deprotonated Glu residues in bacteriorhodopsin) might be advantageous from the perspective of balancing significant binding affinity and rapid proton release, which clearly is of functional importance to biomolecular ion pumps.

Therefore, it can be expected that a proton mediated amino acid pairs exist as an important functional motif, especially in those involved in proton transports. It is worthwhile to emphasize explicitly this type of motif because protons are rarely seen in x-ray structures, although very short distances (< 2.8Å) between heavy atoms are useful indications. Unique IR and NMR features induced by the delocalized proton[459,460] are notable signatures, although, as this study highlights, they need to be carefully analyzed for proper structural interpretation.

Chapter 7

Structure and Characteristics of O State

Summary

The chapter begins with an introduction to the unresolved issues in the characterization of the O structure are described in section 7.1 followed by the details of the simulation setup in section 7.2. The results of the investigation of structural and spectroscopic characteristics of O and [O] states are described in section 7.2.6. PT pathways between O and [O] states were calculated starting from several O models and are summarized in section 7.4. In section 7.5, the results of the excitation energy calculations by including protein polarization effects are explained followed by analysis of retinal geometry in section 7.6.

7.1 Introduction

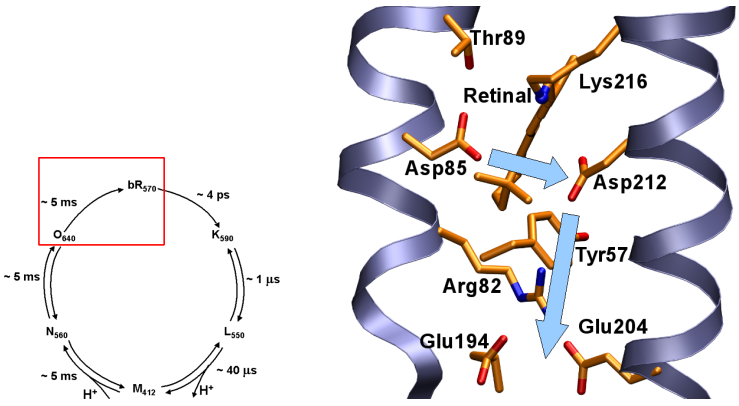
The light-induced proton-pumping cycle of bacteriorhodopsin consists of five sequential proton transfer (PT) steps (Figure 7.1). A proton is first transferred from the retinal's Schiff base to the nearby Asp85. The Asp85 site is electrostatically coupled to an extracellular group (denoted as the proton release group (PRG)),[296,461] which releases a proton to the bulk once Asp85 is protonated. The Schiff base is then reprotonated from the cytoplasmic group Asp96, which will in turn receive a proton from the cytoplasmic bulk (Figure 7.1). The last PT step (the O→bR transition in Figure 7.1a), and the focus of the present work, is from the primary acceptor

Asp85 to the extracellular PRG over a distance of ∼12 Å (Figure 7.1b). The mechanism of the Asp85 deprotonation and the structural details of the conformer ready for the last PT step are poorly understood. Key open questions are: (i) does the PT occur in a single step, from Asp85 to the PRG, or it involves intermediate(s), and (ii) what are the structural elements critical for PT from Asp85 to the PRG. To address these questions, we performed an extensive set of computations of the PT paths, excitation energies, and vibrational spectra.

A recent investigation in which vibrational spectra calculated with various structural models indicated that the PRG likely consists of Glu194 and Glu204 sharing a proton.[434] The combined quantum mechanics/molecular mechanics (QM/MM) simulations[434] demonstrated that the excess proton binds to Glu194/Glu204 instead of the water cluster. The structure with the excess proton shared by the negatively charged Glu194 and Glu204 is stable on the ns time-scale of the QM/MM simulations, and gives rise to a continuum band similar to that observed in previous studies.[220] This role of Glu194/Glu204 as the PRG is consistent with previous interpretations,[218,219] but contradicts a more recent proposal that the excess proton is stored on a water cluster.[222,223]

The path from Asp85 to Glu194/Glu204 is curved, and is interrupted by the side-chain of Arg82 (Figure 7.1b) in the light-adapted bacteriorhodopsin ground state (bR). The coupling between the deprotonation of Asp85 and protein conformational changes is unclear. It has been suggested that the intermediate states of the second half of the photocycle (i.e. late-M, N and O; Figure 7.1a) can be approximated by a single protein conformation.[123] Although several crystal structures exist for M, the assignment of these structures to a specific M substate is difficult. Moreover, it has been questioned whether the crystal environment allows the full extent of protein conformational changes associated with late-M *in vivo*.[173,196] Deprotonation of Asp85 is presumably triggered by the thermal isomerization of the retinal from 13-*cis* back to all-*trans* during the transition between the N and O intermediate states (Figure 7.1a). Experiments have indicated that the retinal back isomerization is coupled to protein structural rearrangements: mutations of amino acid residues close to the retinal such as Leu93 (Figure 7.1b), dramatically delays retinal isomerization by 250-fold.[270]

Rearrangement of water molecules is important for the coupling between Asp85 and the PRG;[125,215] but the details of how the long-distance PT from Asp85 to



(a) (b)

Figure 7.1: The bacteriorhodopsin proton-pumping cycle. (a) Schematic representation of the bacteriorhodopsin photocycle showing intermediate states characterized by their absorption maxima. The last PT steps takes place during O→bR transition from Asp85 to PRG (highlighted in red); (b) Location of protein groups essential for the proton transfer. The arrows indicate the proton transfer (PT) from Asp85 to the PRG during O→bR transition. The picture is prepared using the crystal structure of bR[21] and using the VMD package.[74]

the PRG is coupled to local structural rearrangements of the protein and water molecules are not clear.

Intermediate conformers distinguished by the geometry of the retinal chain and by the protonation states of key protein groups may accumulate during the O→bR step. The rate of the deprotonation of Asp85 is influenced by the local environment around Glu204.[233] The sensitivity of the O→bR transition to temperature and pH[80,462] indicates that this step of the photocycle is rather complex. It has been suggested[233] that the thermal back isomerization of the retinal triggers the formation of a transient O intermediate in which retinal is twisted and Asp85 is deprotonated. The transient O intermediate then converts into the bR after relaxation of the retinal twist. Experiments on the Glu194Gln[126,463] and Glu204Gln[126] mutant phenotypes provided further evidence for the presence of a transient O state (denoted here as [O]) between the O and the bR states. According to ref 1, the transfer of the proton from Asp85 to the PRG may involve an intermediate step [O] in which Asp85 proton had been transferred to Asp212.[126] These experiments on bacteriorhodopsin mutants[126,463] raise the intriguing question of whether the transient [O] intermediate could also occur in the wild-type photocycle. However,

the site-directed mutagenesis data must be interpreted with caution because the coupling between the retinal binding site and the PRG demonstrated by experiments[225] and theory[296] could lead to structural changes of the active site in the mutants, and a PT mechanism different from that in the wild-type.

Because it is difficult to separate O from the previous N intermediate, and the O state does not accumulate in significant amounts in the wild-type photocycle,[464] a crystal structure of the wild-type O state has not been solved. Two putative O-like structures have been proposed: for the chloride pumping Asp85Ser mutant (pdb code:1JV7),[274] and for the acid-blue bacteriorhodopsin (pdb code:1X0I).[275] The assignment of these structural models to an O-like state was based on the main characteristics of the O state - neutral Asp85, optical absorption maximum red-shifted relative to the wild-type bR, and all-*trans* retinal. Compared to the bR,[21] both O-like structures[274,275] exhibit significant conformational differences at the extracellular side of the protein (Figure 7.2a). However, it is unclear to what extent differences in the crystal packing (the space-group is $C222_1$ for the Asp85Ser mutant, and P622 for acid-blue bacteriorhodopsin) contribute to the structural differences among the two models.

A recent EPR study[276] using site-directed spin labels illustrates the specific structural changes (relative to the bR) at the extracellular side of the O state of the Glu204Gln mutant - which is thought to be almost identical to the O state of the wild-type bacteriorhodopsin.[276] The changes include an inward movement of helix D towards the proton channel, and an increase of the distance between helices C and F (Figure 7.2a). The structural characteristics of the O state protein in the purple membrane environment are better reproduced[276] by the acid-blue bacteriorhodopsin[275] than by Asp85Ser.[274] Although the EPR study is a valuable test of the protein conformation in O,[276] it does not provide direct information about essential details of structural elements important for PT.

Structural details of the protein groups close to Asp85 can have a significant effect on the energetics of the Asp85 deprotonation. Previous QM/MM computations demonstrated that the proton affinity of Asp85 is greatly influenced by hydrogen bonding water molecules from the active site.[298] Water molecules hydrogen bonding to Asp85 and Asp212 could thus be an important determinant of the energy difference between the O and the putative [O] intermediates.

Water molecules close to the retinal Schiff base may also be important for the

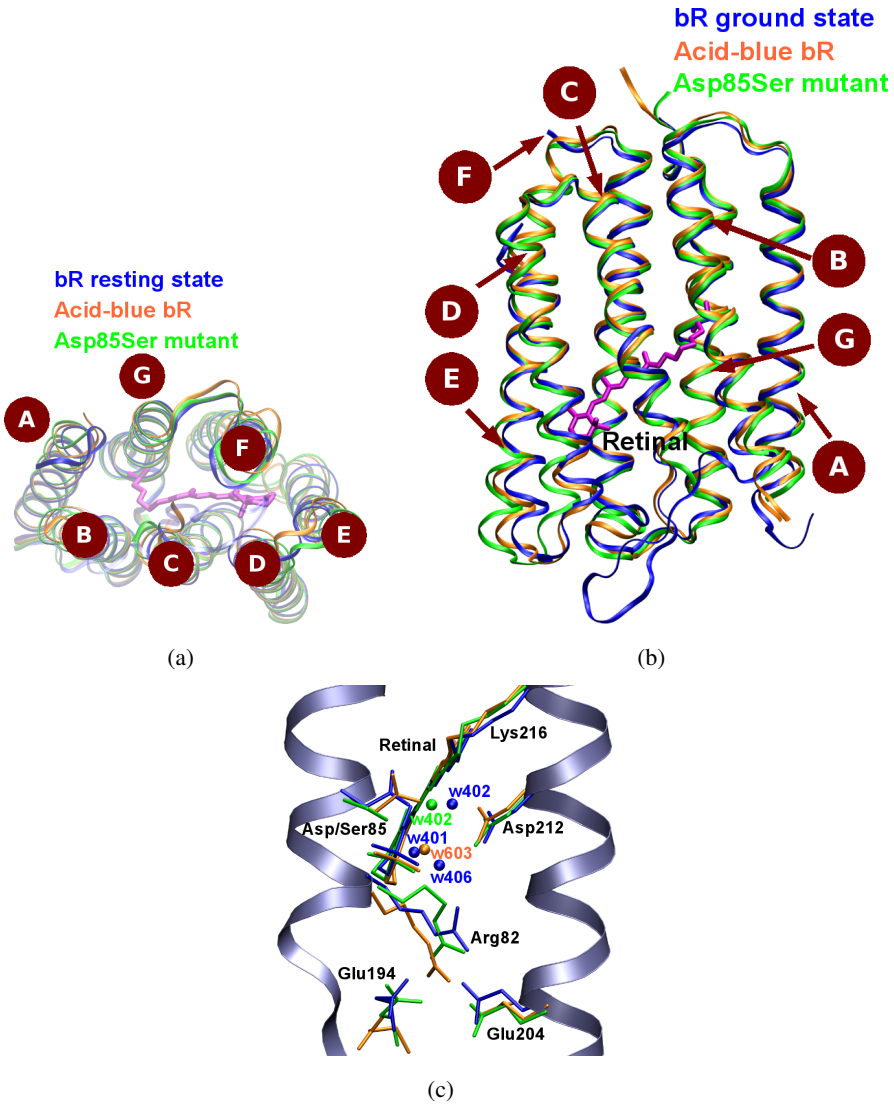


Figure 7.2: Protein conformational changes in O relative to bR. Comparison of the crystal structures of the acid-blue bacteriorhodopsin,[275] the Asp85Ser mutant,[274] and the bR[21] of a view from the cytoplasmic side (a), and parallel to the membrane normal (b). (c) Detailed view of the retinal, water molecules and amino acids important for the last proton-transfer step. The following color codes are used: orange - acid-blue bR; green - Asp85Ser mutant; blue - bR.

structure and spectroscopic properties of the active site. Of the three active site water molecules observed in the bR,[21,76,139,207] only one is present in the O-like

structural models of Asp85Ser mutant or ref 274 and acid-blue bacteriorhodopsin of ref 275. It is unclear whether indeed there is only one water molecule in the O state active site or, the absence of additional water molecules is due to water molecules being disordered.

Due to the proximity between Asp85/Asp212 and the retinal Schiff base (Figure 7.2c), the geometry of the retinal could also influence PT. Resonance Raman studies have suggested that the O state retinal is twisted all-*trans*, with a magnitude of twisting similar to that of the K state structure.[272] The retinal twist relaxes as the photocycle completes with the O→bR transition. Both O-like structures[274,275] have all-*trans* retinal, possibly mixed with structurally similar 13-*cis*, 15-*syn* retinal.[274,275]

Given the electrostatic coupling between the active site and the PRG,[225,296] the geometry of the positively charged Arg82 side-chain may be another important determinant of the energetics of PT. In the bR, the guanidinium group of Arg82 is oriented towards the retinal Schiff base region. In contrast, O-like structural models[274,275] indicate an orientation of the Arg82 side-chain towards the extracellular Glu194/Glu204 (Figure 7.2c). If the guanidinium group of Arg82 indeed adopts an extracellular orientation in O,[274,275] it must then return to its original cytoplasmic orientation towards the Schiff base during the O→bR transition. The orientation of Arg82 could also influence the dynamics of the water molecules visiting transiently the protein interior, and whether or not they can reach the active site. It is not known at what stage during the O→bR transition the side-chain of Arg82 changes its orientation, and whether or not the movement of Arg82 is part of the PT reaction.

In the absence of accurate structural information on the wild-type O intermediate, it is necessary to understand how key structural features affect the deprotonation of Asp85, and to use a comparison between theory and experiments to validate (or invalidate) structural characteristics of the O and the putative [O] intermediates. The structural features assessed here are: (i) the role of the retinal geometry; (ii) the number of active site water molecules, and (iii) the geometry of Arg82 side-chain. We computed reaction paths using seven different protein models. To validate specific structural characteristics of the models, we computed vertical excitation energies and vibrational spectra. The calculations indicate that an [O] intermediate with deprotonated Asp85/protonated Asp212 could indeed be sampled along the

O→bR path. In O state retinal is twisted, and three water molecules are present in the Schiff base region.

7.2 Computational Details

7.2.1 QM/MM setup

The protein is divided in a quantum mechanics (QM) region that contains the groups involved in the chemical reaction and the remaining protein is treated with classical molecular mechanics (MM). The total energy is given by,

$$E = E_{\mathrm{QM}} + E_{\mathrm{MM}} + E_{\mathrm{QM:MM}} \tag{7.1}$$

where E_{QM} is the interaction energy between the electrons and the nuclei of the QM region, E_{MM} describes the interactions between the MM atoms, and $E_{QM:MM}$ represents the interaction energy between the quantum and classical subsystems. The $E_{QM:MM}$ term can be further decomposed in,

$$E_{\mathrm{QM:MM}} = E^{\mathrm{b}}_{\mathrm{QM:MM}} + E^{\mathrm{Coul}}_{\mathrm{QM:MM}} + E^{\mathrm{vdw}}_{\mathrm{QM:MM}} \tag{7.2}$$

where $E^{b}_{QM:MM}$ gives the bonded interactions at the frontier bonds between the QM and MM atoms. $E^{Coul}_{QM:MM} + E^{vdw}_{QM:MM}$ give the Coulombic and Van der Waals interactions respectively between the QM and MM particles.

The interaction energy between the classical particles are calculated using the CHARMM molecular force field.[357] The QM and QM:MM interactions are computed with an approximate DFT method, the self-consistent charge density functional tight binding (SCC-DFTB).[345] The implementation of SCC-DFTB in CHARMM has been described in ref 398. SCC-DFTB is about three orders of magnitude faster as compared to full DFT methods.[348] The accuracy of SCC-DFTB QM/MM approach for studies of biological systems has been documented in refs 348, 398, 448, 465–470. As a prerequisite for the application to retinal proteins, we demonstrated that SCC-DFTB is in good agreement with full DFT methods for describing the ground state properties of protonated Schiff base models (torsional barriers, bond length alternation of the polyene chain etc.).[471] In the context of retinal proteins, SCC-DFTB has been applied to study the primary proton transport in bR,[119,298–300,472] to refine the rhodopsin crystal structure[473] and to generate energy

minimized geometries and molecular dynamics (MD) trajectories for excited states calculations.[321,397,450,474] A detailed description of a QM/MM setup for bacteriorhodopsin is given in ref 298.

The combined QM/MM methods[367,369,370,386] have greatly extended the usefulness of theoretical approaches for understanding complex biological systems. We can calculate spectroscopic properties and investigate the energetic details of chemical reactions. QM/MM methods had been applied to study the dynamics of the initial photo-isomerization event in bacteriorhodopsin,[475,476] the structure and spectroscopic properties of the active site in the bR,[297] and the early intermediate K structure,[472,477,478] and of protonated water clusters in bacteriorhodopsin.[303,304] The mechanism of the first PT event has been investigated in detail,.[119,298–301,472] Excited states properties of bacteriorhodopsin have been studied with QM/MM approaches,[321,450,479–482] several of them including polarization effects from the protein environment.[397,474,475,483]

The QM/MM studies above and of systems such as carbonic anhydrase II[405] or Staphylococal nuclease[484] have illustrated the strengths and the limitations of the current QM/MM methods.[370,449,485,486] The QM/MM computations provided an understanding at an unprecedented level of detail of the possible scenarios for the first PT step and of the energetics associated with the PT paths in bacteriorhodopsin.[119,298,300,301,472] However, modern computational models cannot yet reproduce experimental results within the experimental error - e.g., reaction energies within 1 kcal/mol, vibrational frequencies within a few cm^{-1}, and excitation energies within few nanometers. In spite of the limitations regarding the computation of absolute values, QM/MM computations with a proper treatment of the QM region are reliable for the investigation of relative values.

7.2.2 Proton Transfer (PT) Pathways

To investigate the minimum-energy pathways, we used the Conjugate Peak Refinement (CPR)[487] method as implemented in the TRAVEL module of CHARMM. The CPR method does not require an explicit definition of the reaction coordinate, but only the energy-minimized reactant and product structures. We investigated the PT scenario in all the structural models described in section 7.2.6 below. The intermediate path points between the saddle points of the converged CPR pathway were further refined using the Synchronous Chain Minimization (SCM) algorithm[488]

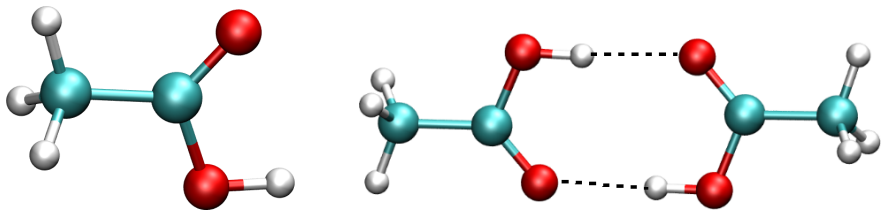
implemented in the TREK module of CHARMM.

7.2.3 Absorption Shifts

We calculated the excitation energies using a high-level *ab initio* method, Spectroscopy Oriented Configuration Interaction (SORCI)[320] as implemented in the ORCA quantum chemical package.[489] SORCI utilizes the division of the first-order interaction space into weakly and strongly perturbing configurations, thereby combining the concepts of the classical multireference configuration interaction and the multireference perturbation theory. While the strongly perturbing configurations are treated variationally, the weakly perturbing configurations are treated with second-order Møller-Plesset perturbation theory. The use of approximate natural orbitals eliminates the problem of choosing a suitable single-particle basis whose quality would affect the final CI result. SORCI gains computational efficiency by use of several thresholds which have been carefully adjusted for the system under study. The thresholds used are: $T_{\mathrm{Pre}} = 10^{-3}$, $T_{\mathrm{Nat}} = 10^{-6}$ and $T_{\mathrm{Sel}} = 10^{-6} E_{\mathrm{h}}$ (see supporting information of ref 321). Only core orbitals are frozen. We used Ahlrich's SV(P) basis set[322] which is appropriate for the calculations of the complete retinal. SORCI has been successfully used to calculate vertical excitation energies for retinal proteins previously, based on optimized geometries calculated with SCC-DFTB/CHARMM.[321] The spectral shifts between proteins are described very well e.g. the shift between bR and SRII could be explained on the basis of structural determinants,[450] but the absolute excitation energies are overestimated by approximately 0.15 eV.[321] The overestimation of the excitation energies is partly due to the effects of polarization and dispersion,[397,474] which are not covered in the widely applied standard QM/MM methods. Since calculations that account for these effects are rather costly, we perform standard QM/MM excited states calculations for all the models and recalculate the excitation energies with polarization effects for several selected models. The effect of protein polarization is described using an interactive polarizable model in which we assign atomic polarizabilities to MM atoms of the protein in addition to the fixed atomic charges.[397,474]

7.2.4 Vibrational Frequencies

The standard normal mode analysis requires diagonalization of the calculated mass-weighted second derivative matrix (Hessian). Since for a system size of several thousand atoms like bacteriorhodopsin, diagonalization of the full Hessian matrix is computationally demanding, we perform the diagonalization in a reduced basis. To calculate the vibrational frequencies we used the VIBRAN module[357] of the CHARMM in the QM/MM setup.[490] The advantage of such an approach consists in accurate description of the polarization effect by the surrounding MM environment on the QM region.[490] We computed the vibrational frequencies using SCC-DFTB for the QM region. Recent studies have shown that SCC-DFTB is accurate in reproducing DFT results for vibrational frequencies of various systems.[447,491]



(a) (b)

Figure 7.3: Gas Phase models for carboxylate using acetic acid. (a) An isolated molecule in the protonated state; (b) Two acetic acid molecules forming cyclic dimer with strong hydrogen bonds.

Table 7.1: Symmetric C=O stretch frequencies (cm^{-1}) of gas-phase models

Model	Calculated value (cm^{-1})	Experimental value (cm^{-1})[a]	shift (cm^{-1})
Acetic acid(aa)	1778	1780	2
Cyclic aa dimer	1708	1725	17

[a] For review see[258]

We carried out additional benchmark calculations for the ν_{COOH} band of isolated acetic acid molecule, and for strongly hydrogen bonded cyclic acetic acid dimer (Figure 7.3 and Table 7.1). The calculations demonstrate that SCC-DFTB can accurately reproduce the ν_{COOH} band for the isolated acetic acid molecule (1778 cm^{-1}) compared with the experimental value of 1780 cm^{-1}. For the strongly hydrogen bonded cyclic dimer, the ν_{COOH} band computed with SCC-DFTB is red-

shifted by approximately 17 cm^{-1} relative to the experimental value of 1725 cm^{-1} (Table 7.1).

7.2.5 Molecular Dynamics Simulations

The MD simulations (classical and QM/MM) were initiated starting from the crystal structure of acid-blue bacteriorhodopsin.[275] Equilibrium simulations with stochastic boundary condition and explicit solvent were carried out using CHARMM package (version: c32a2).[357] In all cases, the system was partitioned into a 22 Å spherical inner region around N_ε atom of Arg82 while the remaining region was treated as outer region following the GSBP approach.[405,426,449] Newtonian equations-of-motion were solved for the molecular dynamics region (within 18 Å), and Langevin equations-of-motion were solved for the buffer region (18-22 Å) with a temperature bath of 300 K..[492] Protein atoms at the boundary between inner and outer regions are constrained following previously described protocol.[426] For classical MD simulations (3 independent 5-ns trajectories), all bonds involving hydrogen for were constrained using SHAKE[493] to allow a 1-fs timestep. For QM/MM MD simulations (3 independent 1-ns trajectories), QM level is SCC-DFTB[345] and the third-order extension of SCC-DFTB[447] was used which was found to be important for improving the proton affinity.[447] The QM region consisted of complete retinal, Asp85, Asp212, Lys216 side chain and 3 water molecules (w401*, w603, and w406*). All bonds involving hydrogen atoms outside the QM region were constrained to allow 1-fs timestep.

7.2.6 Structural Models

The results reported here are based on three structural models (model-A, model-B and model-C; see below) distinguished by the protein structure, the geometry of retinal and the Arg82 side-chain, and by the number and location of water molecules close to Asp85/Asp212 (Table 7.2). To investigate in more detail the role of water molecules, we performed four additional sets of computations in which one or two water molecules not present in the crystal structure[274,275] were added in the model-A and model-B structures (Table 7.2). The computations using model-A and model-B with additional water molecules were pursuant to our preliminary observation that the presence of three water molecules in the active site of model-A

and model-B leads to an active site geometry similar to that of the bR. For all models used, we denote an O state as neutral Asp85 and negatively charged Asp212 and a [O] state as negatively charged Asp85 and neutral Asp212.

Models with Acid-blue form of bR (model-A). To prepare the model-A we used the crystal structure of acid-blue bacteriorhodopsin.[275] This structure indicates coordinates for one active site water molecule (w603) that bridges Asp85 and Asp212 (Figure 7.2c). We prepared the following models: (i) model-A^{1w} contains the single active site water molecule w603 from the crystal structure (Figure 7.4a-b); (ii) model-A^{2w} contains w603 from the crystal structure and w401* (Figure 7.5a-b); (iii) model-A^{3w} contains water molecules w603, w401* and w406* (Figure 7.8a-b).

Models with Asp85Ser mutant (model-B). The model-B structural conformers were prepared starting from the Asp85Ser mutant crystal structure.[274] The crystal structure has a single water molecule w402 in the active site, and w402 is hydrogen bonded to the Schiff base (N-O distance: 2.7 Å). We mutated Ser85 back into Asp85 and prepared end-conformers for: model-B^{1w} (one active site water molecule w402; Figure 7.4c-d), model-B^{2w} (two active site water molecules w402 and w401*; Figure 7.5c-d), model-B^{3w} (three active site waters molecules w402, w401* and w406*; Figure 7.8c-d).

Model with bR (model-C^{3w}). To assess how the PT energetics is influenced by the protein structure, the orientation of the Arg82 side-chain and by the active site water molecules, we performed an additional set of computations starting from the coordinates of the bR (pdb code: 1C3W).[21] The active site of model-C^{3w} contains three water molecules: w401, w402 and w406 (Figure 7.9a-b).

7.2.7 System Setup

Asp96 and Asp115 are modeled neutral as indicated by experiments.[126,273] Standard protonation states are used for all remaining titratable protein amino acids. All buried water molecules present in the crystal structures are included in the calculations. The hydrogen atoms are built using the HBUILD utility of CHARMM package.[357]

The QM region consisted of 78 atoms (for model-A^{1w} and model-B^{1w} with one active site water molecule), 81 atoms (for the model-A^{2w} and model-B^{2w} models with 2 active site water molecules), or 84 atoms (for the models with 3 active

Table 7.2: Details of the end states of the various O models[a]

Model	Starting crystal structure	Protonation state	Active site water molecules(s)
Model-A^{1w}-O	Acid-blue bR[b]	Asp85^{0}/Asp212^{-1}/Glu204^{-1}	w603
Model-A^{1w}-[O]	Acid-blue bR	Asp85^{-1}/Asp212^{0}/Glu204^{-1}	w603
Model-A^{2w}-O	Acid-blue bR	Asp85^{0}/Asp212^{-1}/Glu204^{-1}	w603, w401*
Model-A^{2w}-[O]	Acid-blue bR	Asp85^{-1}/Asp212^{0}/Glu204^{-1}	w603, w401*
Model-A^{3w}-O	Acid-blue bR	Asp85^{0}/Asp212^{-1}/Glu204^{-1}	w603, w401*, w406*
Model-A^{3w}-[O]	Acid-blue bR	Asp85^{-1}/Asp212^{0}/Glu204^{-1}	w603, w401*, w406*
Model-B^{1w}-O	Asp85Ser mutant[c]	Asp85^{0}/Asp212^{-1}/Glu204^{-1}	w402
Model-B^{1w}-[O]	Asp85Ser mutant	Asp85^{-1}/Asp212^{0}/Glu204^{-1}	w402
Model-B^{2w}-O	Asp85Ser mutant	Asp85^{0}/Asp212^{-1}/Glu204^{-1}	w402, w401*
Model-B^{2w}-[O]	Asp85Ser mutant	Asp85^{-1}/Asp212^{0}/Glu204^{-1}	w402, w401*
Model-B^{3w}-O	Asp85Ser mutant	Asp85^{0}/Asp212^{-1}/Glu204^{-1}	w402, w401*, w406*
Model-B^{3w}-[O]	Asp85Ser mutant	Asp85^{-1}/Asp212^{0}/Glu204^{-1}	w402, w401*, w406*
Model-C^{3w}-O	bR[d]	Asp85^{0}/Asp212^{-1}/Glu204^{-1}	w401, w402, w406
Model-C^{3w}-[O]	bR	Asp85^{-1}/Asp212^{0}/Glu204^{-1}	w401, w402, w406

[a] Extra added waters are marked with asterisk(*); [b] ref 275;
[c] ref 274; [d] ref 21

site water molecules). The QM region comprised Lys216, the complete retinal molecule, the side-chains of Asp85, Asp212, and the active site water molecule(s). The boundary between the quantum and classical subsystems was chosen by adding the link atom at the C_β-C_γ covalent bond for Lys216, and at the C_α-C_β covalent bond of Asp85 and Asp212. The QM/MM frontier was treated using DIV linking scheme.[374] In the DIV scheme, the partial charge of the MM host atom is removed from the host and distributed evenly over the remaining MM host atoms.[374]

To maintain the shape of the protein close to the crystal structure we fixed a part of the protein backbone to the crystal structure coordinates. The remaining part was mobile and consisted of 1578 atoms including the quantum mechanical region and a layer of protein groups and surrounding water molecules that can be related to the PT path. The SCC-DFTB/CHARMM energy minimizations were performed to a RMS energy gradient of 10^{-3} kcal/mol.Å.

7.3 Structural Characteristics of End States

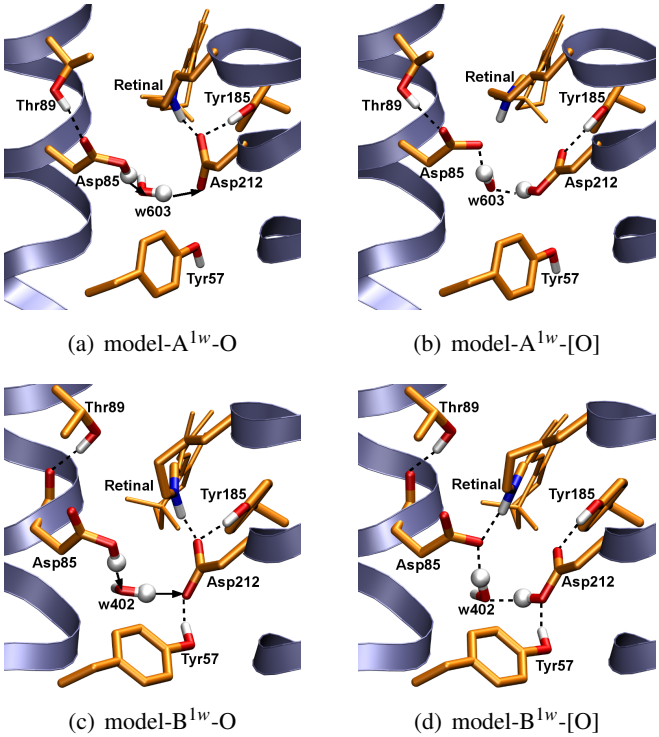
7.3.1 Models with One Active Site Water Molecule

In the crystal structure of the acid-blue bacteriorhodopsin,[275] the distance between Asp85-$O^{\delta 1}$ and Asp212-$O^{\delta 2}$ is shorter (4.2 Å) than in the bR (5.1 Å) of ref 21 The shorter distance between Asp85 and Asp212 in the O state was explained by either Asp85 or Asp212 being protonated. In the QM/MM-optimized end-conformers of model-A^{1w}, retinal is twisted. The Schiff base salt bridges to Asp212 in O (model-A^{1w}-O: Figure 7.4a), and to Asp85 in [O] (model-A^{1w}-[O]: Figure 7.4b).

In the crystal structure of the Asp85Ser mutant[274] water molecule w402 is hydrogen bonded to the Schiff base (O-N distance: 2.7 Å). In the QM/MM-optimized O conformer of model-B^{1w} the Schiff base:w402 hydrogen bond is broken and w402 is displaced towards a more extracellular location, where it forms hydrogen bonds with Asp85 and Asp212 (Figure 7.4c). Retinal is twisted and the Schiff base salt bridges to Asp212 in O (model-B^{1w}-O: Figure 7.4c), and to Asp85 in [O] (model-B^{1w}-[O]: Figure 7.4d). That is, the structures of model-A^{1w} and model-B^{1w} are similar.

The salt bridge between the retinal Schiff base and Asp212 (in O conformer) and between the retinal Schiff base and Asp85 (in [O] conformer) leads to significant blue-shift of the [O]-O excitation energies relative to experimental [O]-O shift 0.04 eV.[126] The absolute excitation energies in model-A^{1w} are 2.09 eV for O, and 2.16 eV for [O]. In model-B^{1w}, the excitation energies are 2.20 eV for O, and 2.44 eV for [O] (Table 7.3). The differences between the excitation energies of model-A^{1w} and model-B^{1w} models can be explained by differences in the hydrogen bonding pattern of the active site. In model-A^{1w} conformers there is a hydrogen bond between Thr89:Asp85, whereas the Tyr57:Asp212 hydrogen bond is absent (Figure 7.4a-b). In contrast, in model-B^{1w} conformers, the Thr89:Asp85 hydrogen bond is not present, but the Tyr57:Asp212 hydrogen bond is present (Figure 7.4c-d). There are also differences in the retinal geometry (see below).

The pattern of hydrogen bonding in the active site may be particularly important for the vibrational frequencies of the ν_{COOH} band. The ν_{COOH} vibrational frequencies of the both model-A^{1w}-O (1690 cm^{-1}) and model-A^{1w}-[O] (1652 cm^{-1}) conformers are largely red-shifted by $\sim$60 cm^{-1} relative to the experimental values (Table 7.4), indicating a more distorted hydrogen bonded network than ex-



(a) model-A^{1w}-O (b) model-A^{1w}-[O]

(c) model-B^{1w}-O (d) model-B^{1w}-[O]

Figure 7.4: Reactant (O-like) and product ([O]-like) states of PT in the presence of a single water molecule in the active site. The Thr89:Asp85 hydrogen bond is present in model-A^{1w}-O (a) and model-A^{1w}-[O] (b) and absent in model-B^{1w}-O (c) and model-B^{1w}-[O] (d). The Tyr57:Asp212 hydrogen bond is absent in model-A^{1w}-O (a) and model-A^{1w}-[O] (b), and is present in model-B^{1w}-O (c) and model-B^{1w}-[O] (d).

pected from experiments. The red-shift in the vibrational frequencies computed with model-A^{1w} can be attributed (at least in part) to the presence of Thr89:Asp85 hydrogen bond. The ν_{COOH} vibrational frequencies computed with model-B^{1w}-O and model-B^{1w}-[O] conformers in which the Thr89:Asp85 hydrogen bond is absent, are within $\sim$10 cm^{-1} from the experimental values (Table 7.4). Moreover, the O-[O] shifts computed with model-A^{1w} (38 cm^{-1}) and model-B^{1w} (39 cm^{-1}) are in good agreement with the experimental value of 42 cm^{-1}.

Table 7.3: S_1 Excitation energies (in eV) and spectral shifts (in eV) for various models[a].

Model	State	Excitation Energy	bR-O shift[b]	Experimental bR-O shift	[O]-O shift	Experimental [O]-O shift
model-A^{1w}	O	2.09	0.41	0.24		
	[O]	2.16			0.07	0.04
model-B^{1w}	O	2.20	0.30	0.24		
	[O]	2.4			0.23	0.04
model-A^{2w}	O	1.92	0.58	0.24		
	[O]	2.06			0.14	0.04
model-B^{2w}	O	2.17	0.33	0.24		
	[O]	2.32			0.16	0.04
model-A^{3w}	O	1.96	0.54	0.24		
	[O]	2.03			0.07	0.04
model-B^{3w}	O	2.15	0.35	0.24		
	[O]	2.29			0.14	0.04
model-C^{3w}	O	1.85	0.65	0.24		
	[O]	1.91			0.06	0.04

[a] Experimental values: br: 2.18 eV,[73] O state: 1.94 eV,[190] [O] state: 1.98 eV[126].

[b] Calculated excitation energy of bR: 2.50 eV (Table 7.6).

7.3.2 Models with Two Active Site Water Molecules

In model-A^{2w} prepared from the acid-blue bacteriorhodopsin structure,[275] we inserted w401* within hydrogen bonding distance from the retinal Schiff base (Figure 7.5a-b). The distance between Asp85-$O^{\delta 1}$ and Asp212-$O^{\delta 2}$ is shorter (4.0 Å) that in the bR (5.1 Å of ref 21). In O conformer of model-A^{2w} (model-A^{2w}-O), w401* hydrogen bonds with the Schiff base (2.7 Å), with the negatively charged Asp212 (2.8 Å), and with Asp85 (2.7 Å: Figure 7.5a). As a result, the presence of the w401* and w402 in the active site reduces the retinal twist relative to models (model-A^{1w} and model-B^{1w}) in which only one water molecule was present (Table 7.7). In [O] conformer of model-A^{2w} (model-A^{2w}-[O]), the hydrogen bond

Table 7.4: C=O stretching band (in cm^{-1}) in various O models[a].

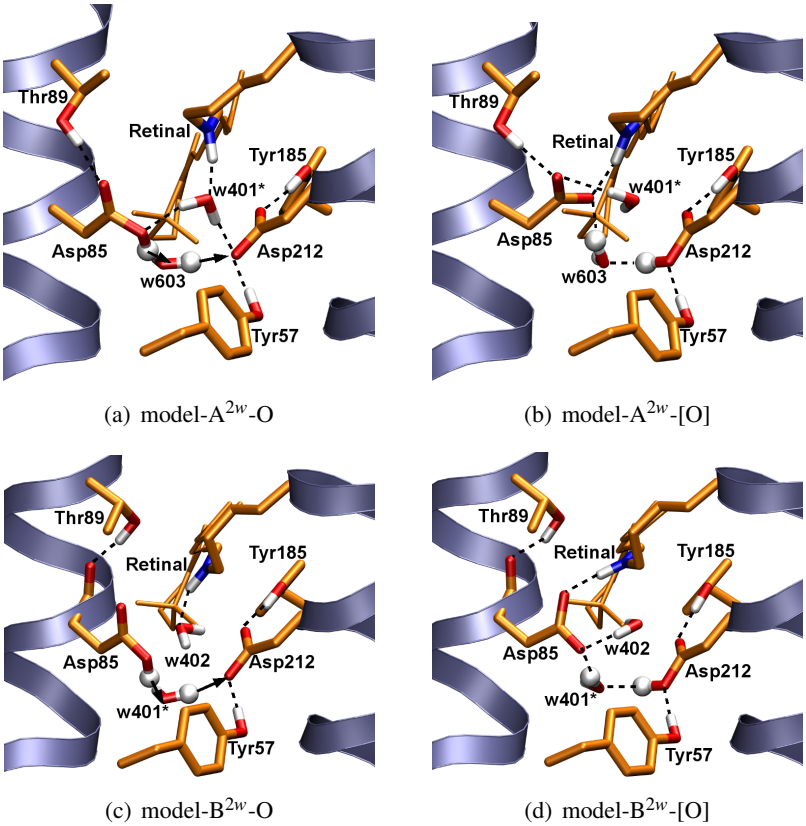
Model	State	IR band	O-[O] shift[b]
Model-A^{1w}	O	1690	38
	[O]	1652	
Model-B^{1w}	O	1741	39
	[O]	1702	
Model-A^{2w}	O	1741	39
	[O]	1702	
Model-B^{2w}	O	1758	54
	[O]	1704	
Model-A^{3w}	O	1737	37
	[O]	1700	
Model-B^{3w}	O	1756	37
	[O]	1719	
Model-C^{3w}	O	1728	24
	[O]	1704	

[a] Experimental values:[126,273,494] Asp85 in O state: 1752-1756 cm^{-1}, Asp212 in transient O([O]): 1712 cm^{-1}.[126,273] [b] Experimental O-[O] shift $\approx$ 42 cm^{-1}.

between the Schiff base and w401* is broken, the retinal chain is twisted, and the Schiff base salt bridges to Asp85 (2.8 Å: Figure 7.5b).

The excitation energy of model-A^{2w}-O (1.92 eV) is close to the experimental value of 1.94 eV, indicating that binding of the retinal Schiff base to water is an important structural determinant of the electronic structure of the retinal. Due to the salt bridge between the retinal Schiff base and Asp85, the absolute excitation energy of model-A^{2w}-[O] state (2.06 eV) is somewhat blue-shifted relative to the experimental value of 1.98 eV. The O-[O] shift (0.14 eV) of model-A^{2w} is much larger than the experimental value of 0.04 eV. Therefore, accurate models of the O and [O] states may be characterized by the retinal Schiff base being hydrogen bonded to water.

Model-B^{2w} was prepared from the Asp85Ser mutant[274] by adding water molecule w401* (Figure 7.5c-d). The O state of model-B^{2w} (model-B^{2w}-O: Figure 7.5c) has



(a) model-A^{2w}-O (b) model-A^{2w}-[O]

(c) model-B^{2w}-O (d) model-B^{2w}-[O]

Figure 7.5: Reactant (O-like) and product ([O]-like) states in the presence of two active site water molecules. The retinal Schiff base hydrogen bonds with w401* in model-A^{2w}-O (a), with w402 in model-B^{2w}-O (c), and with Asp85 in model-A^{2w}-[O] (b) and model-B^{2w}-[O] (d).

similar structural features of the active site as in model-A^{2w}-O: there is a hydrogen bond between the Schiff base and a water molecule (w402 in model-B^{2w}-O and w401* in model-A^{2w}-O) and, a hydrogen bonded chain between Asp85 and Asp212 with the participation of the second active site water molecule (w603 in model-A^{2w}-O and w401* in model-B^{2w}-O).

The higher excitation energy of model-B^{2w}-O (2.17 eV) compared to the excitation energy of model-A^{2w}-O (1.92 eV: Table 7.3) may result from the shorter distance between the Schiff base and the counterion Asp212 (3.4 Å in model-B^{2w}-O, compared to 4.1 Å in model-A^{2w}-O). The shorter distance in model-B^{2w}-O leads to an increased stabilization of the positive charge at the Schiff base in the ground

state and consequently to a larger excitation energy.

In contrast to the models containing a single water molecule in the active site, the presence of two water molecules in model-A^{2w} and model-B^{2w} is associated with the presence of the Tyr57:Asp212 hydrogen bond. The Thr89:Asp85 hydrogen bond is, however absent in the model-B^{2w} conformers (Figure 7.5c-d). The absence of the Thr89:Asp85 hydrogen bond in model-B^{2w}-O leads to a higher ν_{COOH} stretching frequency (1758 cm^{-1}) compared to that computed in model-A^{2w}-O (1741 cm^{-1}), which contains the Thr89:Asp85 hydrogen bond (Figure 7.5a). These observations support the conclusion from the calculations based on models with a single active site water molecule that the Thr89:Asp85 hydrogen bond plays an important role in tuning of the C=O stretch frequency.

7.3.3 Models with Three Active Site Water Molecules

The discussion above indicates that the structural models with one or two water molecules in the active site are not consistent with the experimental spectral fingerprints. The formation of a salt bridge between the retinal Schiff base and either Asp85 or Asp212 leads to a very distorted retinal geometry, and to optical excitation energies that are blue-shifted relative to experimental excitation energies of O and [O] states (Table 7.3).

To further investigate the configuration of the water molecules in the active site, and the relationship between the location of water molecules and the optical excitation spectra, we performed the following two sets of MD simulations. To investigate the dynamics of the active site water molecules on a time-scale longer than accessible with QM/MM MD simulations, in the first set, three MM MD simulations (5ns each) for model-A^{1w}-O were performed. During these MD simulations w603 moved away from Asp85/Asp212, and the Asp85-$O^{\delta 1}$:Asp212-$O^{\delta 2}$ distance decreased to $\sim$2.7 Å (Figure 7.6c), i.e., much shorter than the 4.2 Å distance indicated by the crystal structure.[275] We also observed that water molecules moved from the PRG region closer to the active site (Figure 7.6b). Therefore one cannot exclude that in O there may be more water molecules in the active site than indicated by the crystal structure.

In the second set, QM/MM MD simulations for model-A^{3w}-O were performed (Figure 7.7). It was observed that the water molecule w401* that connects the Schiff base to Asp85/Asp212 remained in the active site during the simulations

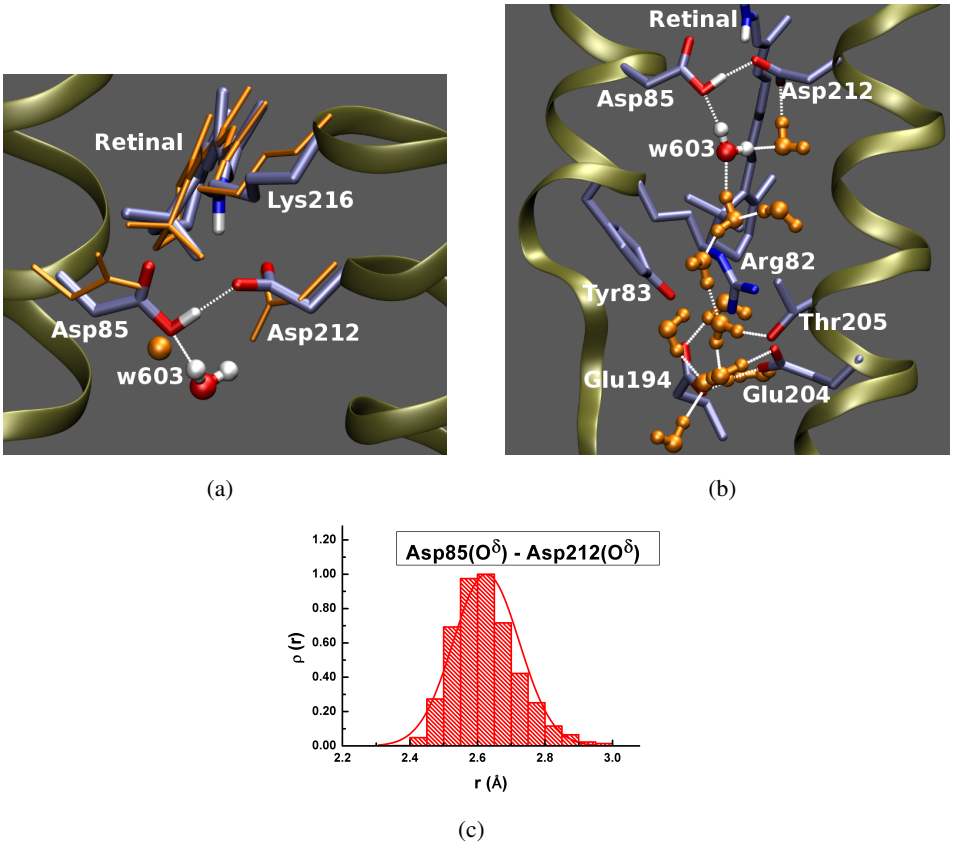


Figure 7.6: Classical molecular dynamics simulations of model-A^{1w}-O. (a) Comparison of the active site of the crystal structure[275](orange) with a snapshot from the simulation shows that the essential structural details of the crystal structure are not preserved during the MD simulations; (b) A hydrogen bonded network forms that connects the proton donor group Asp85 with the PRG Glu194/Glu204; (c) The distance between Asp85 and Asp212 decreases dramatically from 4.2 Å in the crystal structure[275] to ∼2.7 Å in the simulations.

(Figure 7.7a-b). The Asp85-O$^{\delta 1}$:Asp212-O$^{\delta 2}$ distance (∼4.2 Å; Figure 7.7c) agrees well with the crystal structure[275] value of 4.2 Å. The good agreement between the location of water molecule w603 in the crystal structure[275] (Figure 7.7a) and in the QM/MM MD the simulations of model-A^{3w}-O suggests that three water molecules could be accommodated in the crystal structure. Pursuant to these considerations, computations on models with three water molecules as described below were carried out.

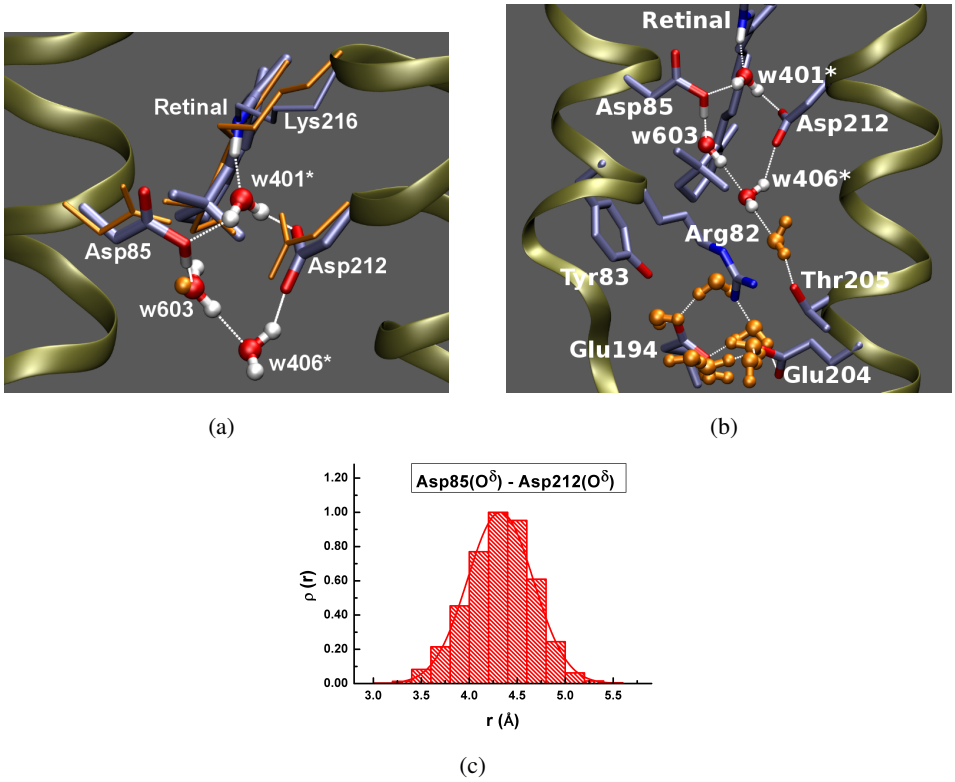


Figure 7.7: The geometry of the active site indicated in ref 275 is well preserved in MD simulation when three water molecules are present in the active site. The good agreement between the crystal structure and the MD simulations with three water molecules (model-A^{3w}-O) indicate that three water molecules are likely present in the active site. (a) Comparison of the active site geometry of the crystal structure (orange) with a snapshot from the SCC-DFTB MD simulation; (b) A snapshot from MD simulation starting from model-A^{3w}-O; (c) Histogram of the distance between Asp85/Asp212 closest oxygen atoms.

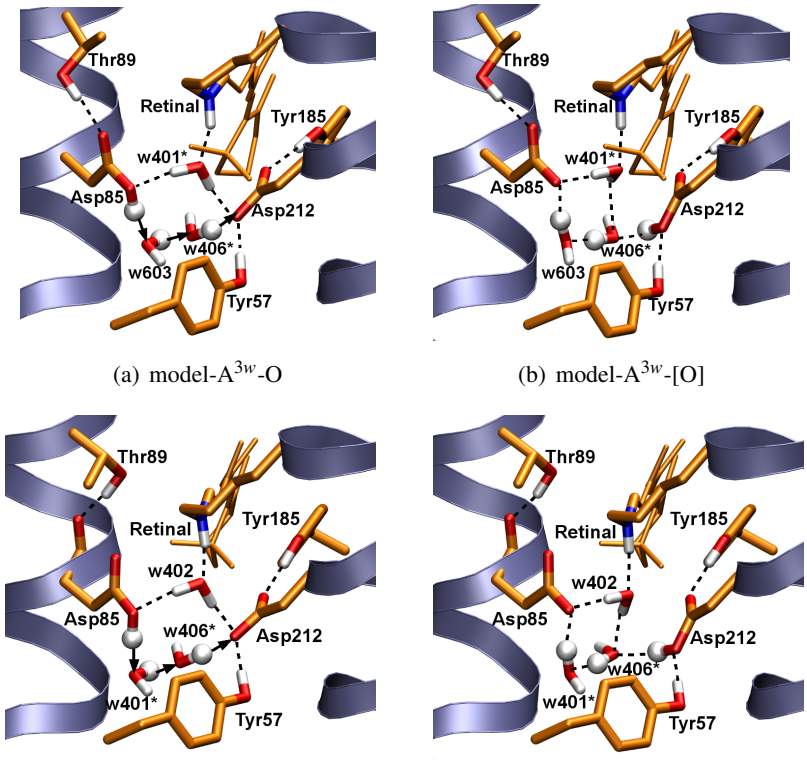
Model-A^{3w}-O (Figure 7.8a) has a hydrogen bonded network in which Asp85 and Asp212 are hydrogen bonded via w603 and w406*; and w401* hydrogen bonds with the Schiff base, Asp85, and Asp212. A similar hydrogen bonded network is observed in model-A^{3w}-[O] (Figure 7.8b). The hydrogen bonding pattern of the active site of model-A^{3w} probably explains the absolute excitation energies of 1.96 eV and 2.03 eV for O and [O], respectively. The calculated excitation energies are close to the experimental values (1.94 for O and 1.98 for [O]). The calculated shift between the [O] and O models (0.07 eV) is in excellent agreement

with the experimental [O]-O shift (Table 7.3).

The hydrogen bonds between Thr89 and Asp85, and between Asp85 and water molecules (Figure 7.8a-b), downshift the ν_{COOH} band in model-A^{3w}-O to 1737 cm^{-1}. In model-A^{3w}-[O] in which Asp212 shares hydrogen bonds with Trp86, Tyr57 and Tyr185, the ν_{COOH} band is at 1700 cm^{-1}. Since SCC-DFTB slightly underestimates these carboxyl stretching frequencies when hydrogen bonds are present (Table 7.1), the frequencies computed with model-A^{3w} are in excellent agreement with the experimental values. The O-[O] shift of the ν_{COOH} band (37 cm^{-1}) is also in very good agreement with the experimental O-[O] shift of 42 cm^{-1} (see Table 7.4).

To estimate the influence of the Thr89:Asp85 hydrogen bond, a model where this hydrogen bond is absent was constructed, and Thr89 hydrogen bonds to the backbone carbonyl oxygen of Asp85. The structural model is 1.1 kcal/mol higher in energy than the model in which the Thr89:Asp85 hydrogen bond is present, and the ν_{COOH} frequency (1758 cm^{-1}) is higher than the experimental value of O state (1752 cm^{-1}).

The active site geometry of the mutant O structures (O and [O] conformers of model-B^{3w}; Figure 7.8c-d) is similar to that of the acid-blue model (Figure 7.8a-b). The main difference between model-A^{3w}-O and model-B^{3w}-O is in the absence of the Thr89:Asp85 hydrogen bond in model-B^{3w}-O (Figure 7.8c). As discussed above for model-A^{3w} conformers without a Thr89:Asp85 hydrogen bond, the absence of this hydrogen bond leads to a blue shifted ν_{COOH} frequency in model-B^{3w}-O. The ν_{COOH} frequency is at 1756 cm^{-1} in model-B^{3w}-O, and at 1719 cm^{-1} in model-B^{3w}-[O]. Given that SCC-DFTB slightly underestimates the C=O frequency in strongly hydrogen bonded structures, the ν_{COOH} frequencies computed for model-B^{3w} conformers are too high relative to the experimental values. Moreover, the excitation energies of 2.15 eV for model-B^{3w}-O and 2.29 for model-B^{3w}-[O] (Table 7.3) are somewhat blue-shifted relative to the experimental values. The blue-shift may be due to the missing Thr89:Asp85 hydrogen bond, and to the short distance between the retinal C_ε atom and Asp85-O$^{\delta 2}$ (3.3 Å in model-B^{3w}-O in contrast to 3.8 Å in model-A^{3w}-O). This supports the proposal that the acid-blue model resembles the conformation of the O state in the purple membrane,[276] and demonstrates that the small spectral shifts calculated between O and [O] for model-A^{3w} are reproducible.



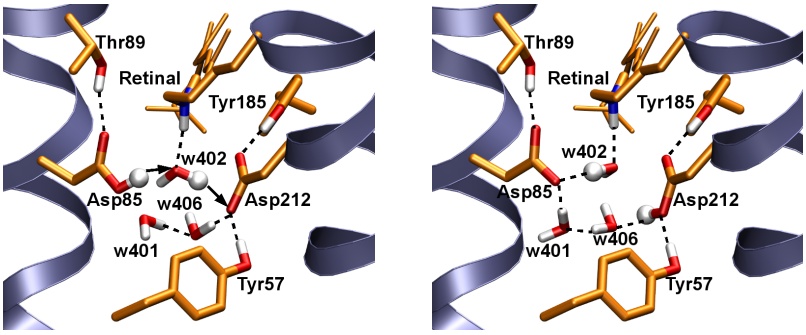
(a) model-A^{3w}-O (b) model-A^{3w}-[O]

(c) model-B^{3w}-O (d) model-B^{3w}-[O]

Figure 7.8: Reactant (O-like) and product ([O]-like) states with three water molecules in the active site. The Schiff base:water hydrogen bond is present regardless of the starting coordinates used for the protein (acid-blue bR^{275} in model-A^{3w} (panels (a) & (b)), and the Asp85Ser mutant in model-B^{3w} (panels (c) & (d)). The Thr89:Asp85 hydrogen bond is present in model-A^{3w} structures (panels (a) & (b)), but absent in model-B^{3w}. The arrows in panels (a) and (c) indicate the PT path.

The bR^{21} has a strong hydrogen bonded network in the active site comprising the Schiff base, the three water molecules w401, w402 and w406, and the two anionic Asp85 and Asp212 groups. The crystal structure of the bR state indicates a cytoplasmic orientation of the Arg82 side-chain,[21] which is preserved in the end states of model-C^{3w} prepared from the bR.[21] Both model-C^{3w}-O (Figure 7.8a) and model-C^{3w}-[O] (Figure 7.8b) have a twisted all-*trans* retinal. The Thr89:Asp85 hydrogen bond present in the bR crystal structure (2.7 Å) is present in both model-C^{3w}-O and model-C^{3w}-[O] conformers. In model-C^{3w}-O, all water molecules from the active site are hydrogen bonded to Asp212; while in model-C^{3w}-[O], the wa-

ter molecules are oriented towards the donor Asp85, and not hydrogen bonded to Asp212.



(a) model-C^{3w}-O (b) model-C^{3w}-[O]

Figure 7.9: Reactant (O-like) and product ([O]-like) states derived from the bR. In the both model-C^{3w}-O (panel (a)) and model-C^{3w}-[O] (panel (b)) the Schiff base hydrogen bonds with w402, and Thr89 hydrogen bonds with Asp85. The arrows in panel (a) indicate the PT path.

In contrast to model-A^{3w} and model-B^{3w}, the absolute excitation energies of model-C^{3w}-O (1.85 eV) and model-C^{3w}-[O] (1.91 eV) are red-shifted compared to the experimental values: 1.94 eV for O (by 0.09 eV) and 1.98 eV for [O] (by 0.07 eV) respectively. Qualitatively, this may be expected due to the cytoplasmic orientation of the guanidinium group of Arg82 in model-C^{3w}-O. In model-C^{3w}-O the positive charge of Arg82 is close to the Schiff base, and thus destabilizes the ground state and decreases the excitation energy relative to ground state. Although the absolute energies of O and [O] states are red-shifted relative to the experimental values (see Table 7.3), the [O]-O shift (0.06 eV) is very close to the experimental value of 0.04 eV.

7.4 PT Pathways

7.4.1 PT in Models with One Active Site Water Molecule

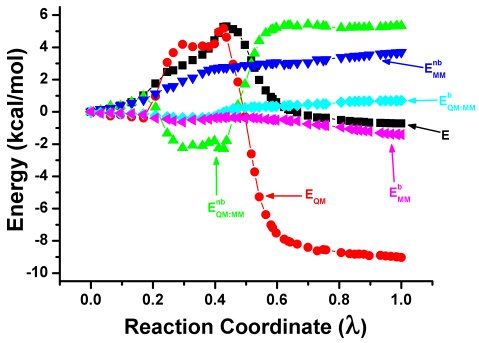
PT in model-A^{1w}:

In model-A^{1w}, PT begins with the reorientation of the Schiff base away from Asp212. As a result, the salt bridge between the Schiff base and Asp212 is weakened. Simultaneously, Asp85 moves closer to w603, such that the Asp85:w603

hydrogen bond is strengthened. At the transition state (λ=0.44 in Figure 7.15a), the hydrogen bonding distances between Asp85:w603 and between w603:Asp212 are equal to 2.4 Å. After the transition state, the retinal chain becomes more planar, and the reaction continues with the transfer of the proton from Asp85 to w603, and from w603 to Asp212, followed by the formation of the salt bridge between the Schiff base and Asp85.

Table 7.5: Activation energies (ΔE_A, kcal/mol) and reaction energies (ΔE_R, kcal/mol) for PT from Asp85 to Asp212

Pathway	Model	Active site water molecule(s)	Member of the Proton Wire	ΔE_A (kcal/mol)	ΔE_R (kcal/mol)
1a	model-A^{1w}	w603	w603	5.3	-0.7
1b	model-B^{1w}	w402	w402	4.8	3.1
2a	model-A^{2w}	w603, w401*	w603	3.9	0.5
2b	model-B^{2w}	w402,w401*	w401*	10.2	5.8
3a	model-A^{3w}	w603, w401*, w406*	w603-w406*	10.9	7.4
3b	model-B^{2w}	w402, w401*, w406*	w401*-w406*	6.7	3.8
3c	model-C^{3w}	w401, w402, w406	w402	13.8	3.6

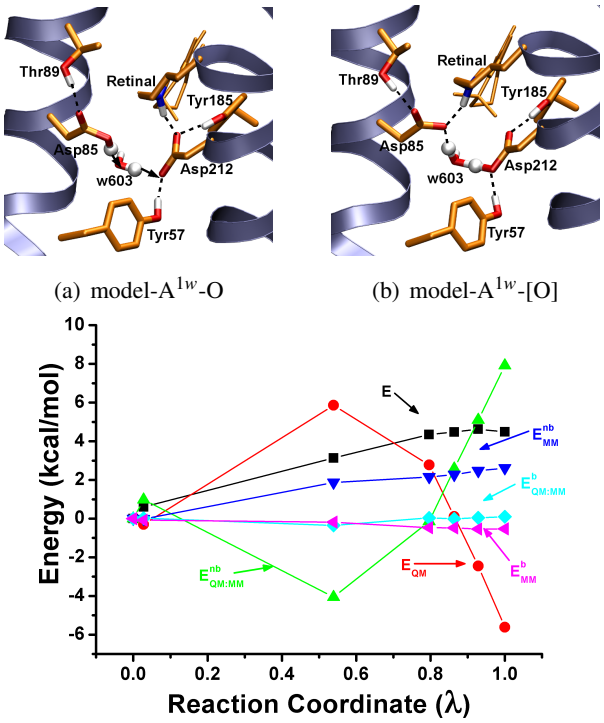


(a) model-A^{1w}

Figure 7.10: Energy profile of PT from Asp85 to Asp212 calculated with model-A^{1w}. The following color codes are used: E (black); E_{QM} (red); $E^{nb}_{QM:MM}$ (green); $E^{b}_{QM:MM}$ (cyan); E^{nb}_{MM} (blue) and E^{b}_{MM} (magenta). At $\lambda = 0$ the proton is on Asp85, and at $\lambda = 1.0$ the proton is on Asp212.

The rate-limiting barrier of 5.3 kcal/mol (λ=0.44; Figure 7.10a) is dominated by E_{QM} (4.6 kcal/mol), and corresponds to the twisted retinal geometry (Figure 7.10a). The favorable contribution of $E^{\mathrm{nb}}_{\mathrm{QM:MM}}$ (-1.79 kcal/mol) to the rate-limiting barrier corresponds to w603 being strongly hydrogen bonded to Asp85 and Asp212. Following the transition state, decrease in the retinal twist is associated with a decrease $\mathrm{E_{QM}}$ energy contribution. The $E^{\mathrm{nb}}_{\mathrm{QM:MM}}$ contribution starts rising from -1.79 kcal/mol at the transition state to 5.3 kcal/mol at $\lambda = 1$, and partially compensates for the negative E_{QM} contribution and leads to an overall exothermic energy profile.

PT in model-A^{1w} in presence of Tyr57:Asp212 hydrogen bond:



(a) model-A^{1w}-O (b) model-A^{1w}-[O]

(c) model-A^{1w}-[O]

Figure 7.11: End states of model-A^{1w} in presence of Tyr57:Asp212 hydrogen bond: (a) O-like; (b) [O]-like; (c) Energy decomposition of minimum energy profile for PT in model-A^{1w} in presence of Tyr57:Asp212 hydrogen bond.

To investigate the role of Tyr57:Asp212 on PT, another set of path calculations were carried out for model-A^{1w} with Tyr57:Asp212 (Fig. 7.11). In presence of the Tyr57:Asp212 hydrogen bond the Asp85:Asp212 distance reduces from 4.2 Å in the crystal structure to 3.7 Å in model-A^{1w}-O conformer (Figure 7.11a). The vibrational calculations indicate for the O conformer, the ν_{COOH} band (1701 cm^{-1}) is nearly same as the ν_{COOH} band in model-A^{1w} without Tyr57:Asp212 hydrogen bond (1690 cm^{-1}), while in [O] conformer as a result of shared proton between w603 and Asp212, the ν_{COOH} band is strongly red-shifted to 1551 cm^{-1}. In model-A^{1w}-[O], the proton is shared between w603 and Asp212 (Figure 7.11b). The PT from Asp85 to Asp212 via w603 is endothermic and the rate-limiting energy barrier of 4.6 kcal/mol (at λ=0.67 in Figure 7.11c) is dominated by energy contribution from $E^{\mathrm{nb}}_{\mathrm{QM:MM}}$ (5.1 kcal/mol) due to the presence of the Tyr57:Asp212 hydrogen bond.

PT in model-B^{1w}:

In the O conformer, the Schiff base forms a salt-bridge with one of the carbonyl oxygen of the acceptor Asp212. The other carbonyl oxygen of Asp212 forms a hydrogen bonded network and involves water w402 and donor Asp85 residue. The reaction begins with change in orientation of the Schiff base which weakens the salt-bridge between the Schiff base and Asp212 residue. Concomitantly, the hydrogen bond between the donor group and w402 strengthens as Asp85 moves closer to the w402. At the transition state (at λ=0.67) (Figure 7.12), the w402 is equally shared between the Asp85 (2.4 Å) and acceptor Asp212 (2.4 Å) groups following which, the proton is completely transferred from Asp85 to w402 and from w402 to Asp212. The reaction proceeds towards completion with formation of salt-bridge between the Schiff base and the Asp85 residue.

Similar to the results for model-A^{1w}, in model-B^{1w} the PT energy barrier is small ($\sim$5 kcal/mol; Table 7.5). Within the accuracy of the methods used, the reaction energy is, however, different in model-A^{1w} and model-B^{1w}, being negative for model-A^{1w}, and positive for model-B^{1w}. That is, model-A^{1w} favors the proton on Asp212, whereas model-B^{1w} favors the proton on Asp85. The different reaction energies computed for model-A^{1w} and model-B^{1w} may be related to the Tyr57:Asp212 hydrogen bond, which is present in model-B^{1w}, but absent in model-A^{1w}. Test computations with a model-A^{1w} conformer in which the Tyr57:Asp212

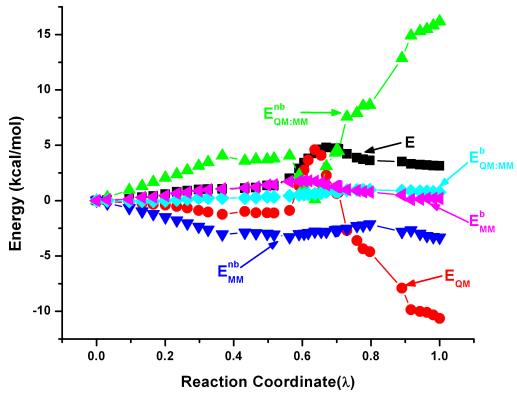


Figure 7.12: Energy decomposition of minimum energy profile for PT in model-B^{1w}

hydrogen bond is present indicate a reaction energy of 4.5 kcal/mol similar to the value of 3.1 kcal/mol computed for model-B^{1w}. It is, however, unlikely that either of the model-A^{1w} and model-B^{1w} geometries are a good representation of the active site geometry in O.

7.4.2 PT in Models with Two Active Site Water Molecules

The mechanism of PT in models with two active site water molecules is similar to that discussed above for models with one active site water molecule. In the case where two active site water molecules are present, the proton can be transferred either via the water molecule hydrogen bonded to the retinal Schiff base, or via the water molecule with the more extracellular location in the active site. When only one water molecule in present in the active site, it acts as an intermediate carrier for the proton.

PT in model-A^{2w}:

In model-A^{2w}, the PT takes place from the donor Asp85 residue via water w603 to the acceptor Asp212 residue. In pathway begins as Asp85 while maintaining its hydrogen bond with Thr89 (2.8 Å) orients such that the distance between proton carrying oxygen of Asp85 (O$^{\delta 1}$) and the Schiff base nitrogen reduces from 4.6 Å to 2.9 Å at the transition state (at λ=0.72) (see Figure 7.13). The rate-limiting barrier (3.92 kcal/mol) is dominated by opposite sign contributions from E_{QM}(-2.2

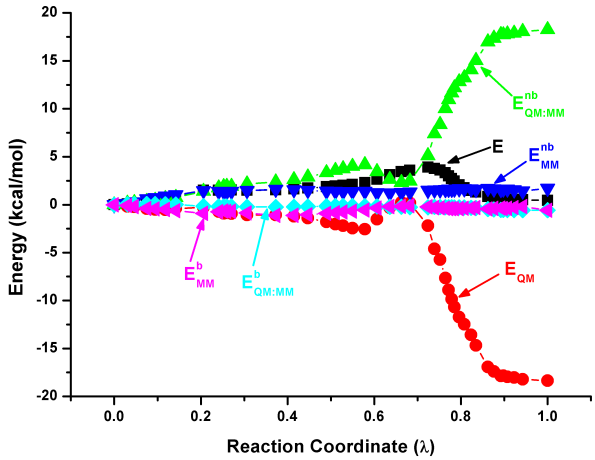


Figure 7.13: Energy decomposition of minimum energy profile for PT in model-A^{2w}

kcal/mol) and $E^{\text{nb}}_{\text{QM:MM}}$(5.1 kcal/mol) and correspond to a shared proton between Asp85/w603 and another proton shared between w603/Asp212. After the PT, a salt bridge is formed between Asp85 and the Schiff base and correspond to almost equal opposite sign contributions from favorable E_{QM} (-18.4 kcal/mol) and unfavorable $E^{\text{nb}}_{\text{QM:MM}}$ (18.3 kcal/mol).

PT in model-B^{2w}:

The energy profile for the proton transfer pathway in the model-B^{2w} model is shown in Figure 7.14. In the beginning of the pathway in model-B^{2w} the hydrogen bond between the Schiff base and w402 breaks as the Schiff base orients away from w402 and towards Asp85. This movement continues till λ=0.64 and involves simultaneous movement of Asp212 such that the distance between the Schiff base and the Asp212-O$^{\delta 2}$ atom increases from 3.4 Å to 4.6 Å. Such increase in the distance between the cationic Schiff base and the anionic Asp212 residue increases the E_{QM} up to 10.5 kcal/mol and decreases the $E^{\text{nb}}_{\text{QM:MM}}$ by 8.4 kcal/mol. The water w402 then orients to form a hydrogen bond with Asp85 residue. Concomitantly, the proton donor Asp85 moves closer the water W401* such that at the transition state at λ=0.68, the Asp85 shares its proton with water W401* and W401* shares

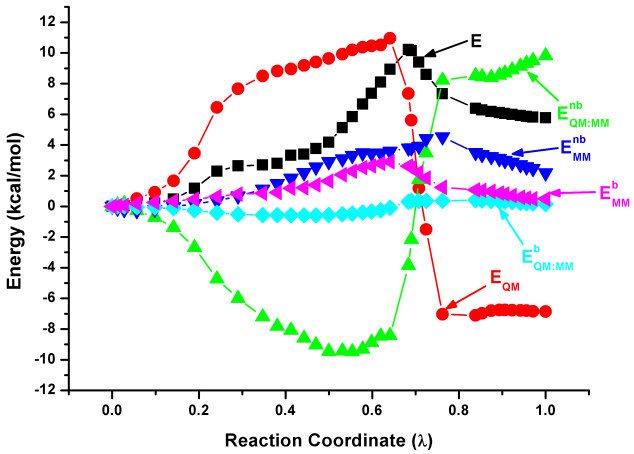


Figure 7.14: Energy decomposition of minimum energy profile for PT in model-B^{2w}

one of its proton with Asp212. Following this, the PT is completed in sequential manner.

7.4.3 PT in Models with three Active Site Water Molecules

PT in model-A^{3w}:

The PT pathway from Asp85 to Asp212 via water molecules w603 and w406* occurs via a concerted mechanism in model-A^{3w} (Figure 7.15b). The pathway starts with the decrease of the distance between Asp212 and w406* from 2.7 Å to 2.6 Å, such that the distances between Asp85 and w603, between w603 and w406*, and between w406* and Asp212 become equal to 2.6 Å. At the transition state (λ=0.47 in Figure 7.15b), protons are shared betwen Asp85-w603 (2.5 Å), between w603-w406* (2.5 Å), and between w406*-Asp212 (2.4 Å). This gives the rate-limiting energy barrier of 10.9 kcal/mol. The reaction then proceeds with PT from Asp85 to w603, from w603 to w406*, and from w406* to Asp212. As a consequence of the transfer, the hydrogen bond between w401* and Asp212 is broken, and a new hydrogen bond is formed between the w401* and w406*. The E_{QM} and the $E^{\mathrm{nb}}_{\mathrm{QM:MM}}$ terms contribute 4.9 kcal/mol and, 5.0 kcal/mol respectively to the rate-limiting

barrier (Figure 7.15b). The increase in $E^{nb}_{QM:MM}$ from 5.0 kcal/mol at the transition state to 13.5 kcal/mol at the end of PT (Figure 7.15b) is due to the increase of the Tyr185:Asp212 and Tyr57:Asp212 hydrogen bonding distances from 2.6Å in the O, to 2.7 Å in [O] state.

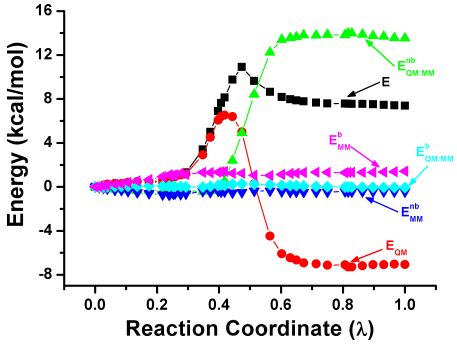


Figure 7.15: Energy profile of PT from Asp85 to Asp212 calculated with and model-A^{3w}.

PT in model-B^{3w}:

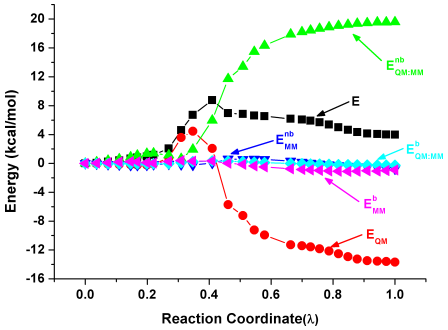


Figure 7.16: Energy decomposition of minimum energy profile for PT in model-B^{3w}

In the O state of model-B^{3w}, the distances between the active site residues Asp85/w401* (2.6 Å), w401*/w406* (2.7 Å) and w406*/Asp212 (2.7 Å) are well optimized for the PT. Thus, the PT reaction proceeds without significant structural change upto λ=0.2 (see Figure 7.16) after which, the distance between Asp85/w401*, w401*/w406* and w406*/Asp212 decreases such that at the transition state (at λ=0.41), the proton is shared between Asp85/w401* (2.5 Å), w401*/w406* (2.4

Å) and w406*/Asp212 (2.4 Å) following which the proton is transferred to the Asp212 completing the pathway. In the [O] state, the weak hydrogen bond between Tyr57 and Asp212 (2.7 Å) and Tyr185 and Asp212 (2.7 Å) from 2.6 Å in the reactant is reflected by increase in the $E^{nb}_{QM:MM}$ to 19.6 kcal/mol relative to the reactant. Following the PT in the [O] state, the hydrogen bond between water w402 and Asp212 is broken and a new hydrogen bond between w402 and w406* is formed.

PT in model-C^{3w}:

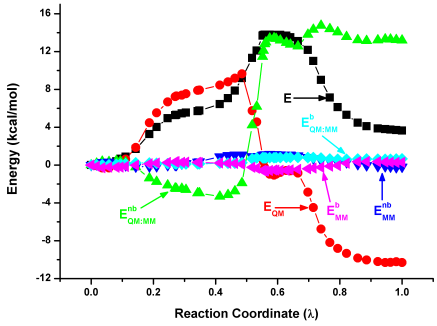


Figure 7.17: Energy decomposition of minimum energy profile for PT in model-C^{3w}

The proton pathway via w402 in model-C^{3w} starts with re-orientation of waters molecules w401 and w406 from Asp212 to Asp85 breaking the hydrogen bond with Asp212. Simultaneously, Asp85 and Asp212 move closer to w402 while w402 maintains its hydrogen bond with Schiff base. Subsequently, at the rate-limiting step (λ=0.56 in Figure 7.17), a proton is shared between Asp85 and w402 (2.5 Å) and between w402 and Asp212 (2.5 Å). The PT then finishes with complete transfer of proton from Asp85 to w402 and from w402 to Asp212. The initial contribution to the total energy is dominated by E_{QM} and is due to the orientation of waters breaking the hydrogen bonds between the waters and Asp212. Once the orientation is complete, $E^{nb}_{QM:MM}$ dominates the contribution to the total energy and contributes mainly (13 kcal/mol) to the rate-limiting energy barrier (13.8 kcal/mol located at λ=0.56). The increase in $E^{nb}_{QM:MM}$ is because of the unfavorable electrostatic interactions (11.8 kcal/mol) between the shared water w402 with Asp85 and

Asp212 and the surrounding MM groups.

PT in model-D^{3w}:

To further investigate the influence of the Arg82 orientation on the PT energetics and spectral fingerprints, model-C^{3w} structure was used to construct a model in which the Arg82 side-chain is oriented towards the extracellular side. This conformer is denoted here as model-D^{3w}. The extracellular-oriented Arg82 side-chain was taken from the late M-state structure (pdb-code: 1CWQ),[181] superimposed onto model-C^{3w}.

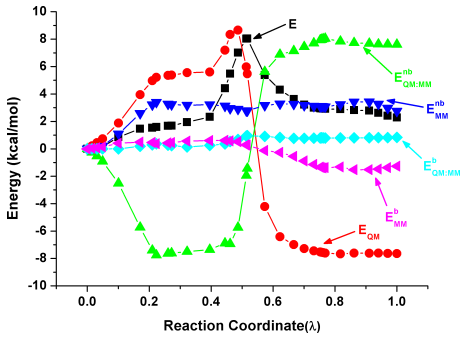


Figure 7.18: Energy decomposition of minimum energy profile for PT in model-D^{3w}

In the O conformer of model-D^{3w}, Thr89 is hydrogen bonded to the backbone carbonyl oxygen of Asp85 and the water w402 to Asp212 and water 406. The pathway begins with orientation of w402 breaking the hydrogen bond with Asp212 and correspond to the increase in E_{QM} (Figure 7.18). Simultaneously, Thr89 orients away from backbone carbonyl Asp85 and forms a hydrogen bond to Asp85 carboxyl oxygen as reflected by sharp decrease in $E^{\mathrm{nb}}_{\mathrm{QM:MM}}$ (at λ=0.26 in Figure 7.18). Following these movements, the proton is transferred in a concerted path from Asp85 to w401, from w401 to w406 and finally from w406 to Asp212. The rate-limiting barrier of 8.0 kcal/mol (at λ=0.52) is dominated by E_{QM} (5.5 kcal/mol) and correspond to the shared proton between Asp85/w401, w401/w406 and w406,Asp212. After the PT, w402 forms hydrogen bond with Asp85 and w406 and correspond to the decrease in E_{QM} while weakening of Tyr57/Asp212 and Tyr185/Asp212 hydrogen bonds from reactant (2.6 Å and 2.6 Å respect.) to [O] (2.7 Å, 2.7 Å respect.) correspond to increase in $E^{\mathrm{nb}}_{\mathrm{QM:MM}}$ at 7.6 kcal/mol. After the

proton is transferred to Asp212, an extracellular water w407 moves to form a hydrogen bond with Thr205 backbone carbonyl oxygen, thus correspond to lowered E^{b}_{MM} energy of -1.3 kcal/mol relative to the reactant.

Thus, the conformer with an extracellular orientation of the Arg82 side-chain is $\sim$4 kcal/mol lower in energy than model-C^{3w}. Due to the larger distance between the Schiff base and the positive charge of Arg82 in model-D^{3w}, the absolute excitation energies of O (2.04 eV) and [O] (2.00 eV) states of model-D^{3w} model are blue-shifted relative to model-C^{3w} (Table 7.3). The rate-limiting PT barrier computed with model-D^{3w} is 8.0 kcal/mol.

7.5 Excited States Calculated by Including Protein Polarization Effects

In a recent study on the influence of the polarizability of the protein/solvent/membrane environment on the calculation of vertical excitation energies, a significant bathochromic shift was found when substituting the fixed CHARMM point-charge model by the polarization model polar.h.[397,474] The latter combines a polarization-free point charge model with an interactive induced-atomic dipole model to treat polarization within the protein environment explicitly. Furthermore, the screening of solvent-exposed charged groups by the solvent/membrane environment may affect the excitation energy. The Poisson - Boltzmann based charge-scaling scheme of Dinner *et al.*[423] represents a simple way to include this effect and is appropriate also for the calculation of excitation energies of retinal proteins.[397]

Table 7.6: Effect of charge scaling (CS) and protein polarization (polar.h) on the excitation energies (eV) of bR and various O Models.

		CHARMM		polar.h
Model	State	no CS	with CS	with CS
bacteriorhodopsin	bR	2.50	2.41	2.24
Model-A^{3w}	O	1.96	1.94	1.92
Model-A^{3w}	[O]	2.03	1.99	2.00
Model-C^{3w}	O	1.85	1.87	1.86

Table 7.6 shows the effects of the charge scaling and the self-consistent polarization of the retinal and its environment employing the polar.h model. In the three O-state models, the net effects are both very small, in contrast to the situation in the bR, where the charge scaling and the polar.h protein electrostatics give rise to bathochromic shifts of 0.09 and 0.17 eV, respectively. The latter is primarily caused by the refined, self-consistent charge distribution in the protein, which includes an electrostatic screening of the doubly charged complex counter ion. This screening reduces both the electrostatic interaction with the retinal and the charge transfer between the protonated Schiff base and the complex counter ion. In the O state, only a single counter ion is present (Asp212), and these effects are less pronounced.

A minor contribution to the larger bathochromic shift in bR originates from the response of the induced dipoles in the protein to the charge transfer associated with the optical excitation of the retinal: When neglecting this response, i.e., keeping the ground-state adapted atomic dipoles fixed, the resulting vertical excitation energies are increased by 0.09 and 0.05 eV for bR and the O state models, respectively. As the polarizability of the protein environment is essentially the same in both cases, the difference correlates with a stronger charge transfer in bR: The difference-dipole moment $|\Delta_{S_1-S_0}\mu|$ is 10.8 debye in bR and 8.4 debye in model-A^{3w}-O conformer.

7.6 Retinal Chromophore Configuration

The crystallographic data has shown that the retinal adopts an almost planar geometry in its isolated form;[495–497] the protein environment induces non-planarity due to steric and electrostatic interactions.[21,67,153,154,498,499] solid-state NMR measurements of the H-C_{14}-C_{15}-H bond along the photocycle indicated that the retinal twist increases by $\sim 17^\circ$ from the bR to early-M.[154] FTIR studies on various mutant bR phenotypes suggested that in the O intermediate the retinal is more twisted than in bR, and that the twist decreases during the O→bR transition.[233]

Resonance Raman studies indicated that in O, the all-*trans* retinal is twisted;[272] the twist is delocalized around the C_7=C_8, C_{11}=C_{12}, and C_{15}=N_{16} bonds.[153,272] The dihedral angles of these retinal double bonds for all O-like models investigated here are given in Table 7.7. The retinal twist is stronger in O than in bR. For

Table 7.7: Torsional angles (in degree) around the double bonds of the polyene chain of retinal in various O models

	C_7=C_8	C_{11}=C_{12}	C_{13}=C_{14}	C_{15}=N_{16}
bR[21]	175.1	178.9	-156.9	-162.7
O Model	C_7=C_8	C_{11}=C_{12}	C_{13}=C_{14}	C_{15}=N_{16}
Model-A^{1w}	176.5	-172.4	-154.6	-155.7
Model-B^{1w}	172.9	-174.4	-168.2	-170.8
Model-A^{2w}	176.2	-171.2	-150.0	-162.9
Model-B^{2w}	173.3	173.8	172.3	178.0
Model-A^{3w}	174.8	-170.9	-151.3	-165.6
Model-B^{3w}	173.2	-175.4	-158.0	-163.8
Model-C^{3w}	170.8	-174.9	-153.4	-162.3

example, the dihedral angle C_{12}-C_{13}=C_{14}-C_{15} is -151.3° in model-A^{3w}-O, and -156.9° in the bR (see Table 7.7). The stronger twist of the retinal chain in the O models than in the bR is consistent with the suggestions from experiments.[233,272] The maximum twists measured for C_6-C_7=C_8-C_9 and C_{10}-C_{11}=C_{12}-C_{13} dihedral angles are $\pm$ 10°. Larger twists of $\pm$ 30° are observed for the C_{10}-C_{11}=C_{12}-C_{13} and C_{14}-C_{15}=N_{16}-C_ε dihedral angles.

Model-A^{1w}-O has large deviations of up to $\sim$20° of the twists for C_{12}-C_{13}=C_{14}-C_{15} and C_{14}-C_{15}=N_{16}-C_ε dihedral angles relative to the planar all-*trans* value of 180° (see Table 7.7). Together with the large blue-shift of the excitation energy and the large red-shift of C=O stretch frequency computed with model-A^{1w}, this distorted retinal geometry indicates that more water molecules may be present in the active site.

Relative to the bR, large change was observed for C_{12}-C_{13}=C_{14}-C_{15} and C_{14}-C_{15}=N_{16}-C_ε dihedral angles for model-A^{3w}-O, model-B^{3w}-O, and model-C^{3w} (Table 7.7). The significant deviations from planarity of the retinal chain in these structural models is consistent with the experimental findings.[233,272]

7.7 Discussion

The aim of this work was to investigate structural models of the O state, and the likelihood that an [O] intermediate characterized by deprotonated Asp85 and protonated Asp212 is sampled during the O→bR transition. To address these questions PT, excitation energies, and vibrational frequencies were carried out using protein models constructed from the crystal structures of acid-blue bacteriorhodopsin,[275] Asp85Ser mutant,[274] and bR.[21] The models used in our computations are distinguished not only by the protein structure, but also by the number of active site water molecules, orientation of the Arg82 side-chain, and by the geometry of the retinal polyene chain.

7.7.1 Role of Active Site Water Molecules

The crystal structures of acid-blue bacteriorhodopsin[275] and Asp85Ser mutant[274] indicate only one water molecule close to the retinal Schiff base. The QM/MM-geometry optimizations reported here indicate that when only one water molecule is present in the active site, local minima in which the Schiff base salt bridges to either Asp85 or Asp212 are energetically preferred. The strong interaction between the Schiff base and Asp85 (or Asp212) leads to optical excitation energies that are blue-shifted relative to the experimental values. That is, the geometries of the active site obtained in the presence of a single water molecule in the active site give spectral fingerprints inconsistent with experiments. Consequently, additional structural models were constructed with two and three water molecules in the active site. The short Asp85:Asp212 distance of 4.2 Å in the acid-blue bacteriorhodopsin crystal structure[275] relative to 5.1 Å in the bR[21] is compatible with three water molecules in the active site, and does not indicate a direct hydrogen bond between Asp85 and Asp212. The mutual repulsion of the negatively charged Asp85 and Asp212 in the bR is likely reduced in O (where Asp85 is neutral). This lesser repulsion in O allows a decrease of the Asp85:Asp212 distance relative to the bR, as indicated in the acid-blue crystal structure.[275] The QM/MM MD simulations on the acid-blue crystal structure (model-A^{3w}-O) indicate that the active site can accommodate three water molecules.

7.7.2 PT Pathways

We calculated PT pathways between the O and [O] states for all investigated models. Since FTIR studies indicate that the protein environment does not change during the O to [O] transition,[463] minimal energy pathways were carried out - i.e., the entropic contribution to the energy was neglected. The entropic contribution is expected to be more important for the PT to the PRG where the proton is transferred over ∼11 Å distance. The activation energy of various PT pathways varies between ∼4.0 kcal/mol and ∼11.0 kcal/mol with the exception of the path in model-C^{3w}. The slightly high barrier for model-C^{3w} (13.8 kcal/mol, path 3c) corresponds to the PT via w402 which is strongly hydrogen bonded to the Schiff base. Except for the path computed from the acid-blue structure with one water molecule (path 1a), the energy profiles of the paths (path 1b, 2a-b, and 3a-c; Table 7.5) indicate small endothermic reaction energies of ∼4 kcal/mol. The reaction energies and PT barriers obtained for the various PT paths support the idea that a transient [O] state with protonated Asp212 can be sampled during the O→bR transition.

7.7.3 UV-Vis Spectral Fingerprints

The calculated optical excitation energies allow us to distinguish between the structural models used. Structural models containing either one or two water molecules in the active site also have a salt bridge between the retinal Schiff base and Asp85 or Asp212. Standard QM/MM computations using these models give excitation energies that are strongly blue-shifted relative to the experimental values. However, since the excitation energy of the bR is also highly overestimated (2.50 eV instead of the experimental value of 2.18 eV), the standard QM/MM calculations are not conclusive.[1] Therefore, the polarization red-shifts was estimated by using a polarizable model for the protein that was developed recently.[397,474] Polarization effects largely red-shift the absorption of the bR to 2.24 eV (see Table 7.6). In contrast, in the case of the O state, the effect of polarization on the excitation energies is very small. From the optical spectra computations it was concluded that the structural models with only one or two water molecules in the active site are unlikely, since they would give a strongly blue-shifted absorption spectrum that is

[1]The absolute excitation energy of bR is 0.07 eV higher than that calculated before at SORCI level(2.34 eV).[321] This difference can be attributed to the large QM region used here (retinal, Asp85, Asp212 and three water molecules) whereas in ref 48, only retinal has been treated with QM.

incompatible with experiments. The Arg82 side-chain is likely oriented towards the extracellular side, because a cytoplasmic orientation red-shifts the absorption spectrum as in case of model-C^{3w}.

The results in Table 7.6 indicate that a proper description of the protein electrostatics is required for a quantitative description of the excitation energies for bR, and of the shift between the bR and the O state. The theoretical value of 0.54 eV for bR-O shift (model-A^{3w}) is still overestimated (compared to the experimental bR-O shift of 0.24 eV). Although the absolute excitation energies are within 0.06 eV of the experimental absorption maxima. The hypsochromic shift from the O to the transient [O] state is reproduced regardless of the protein electrostatic model employed. The excitation energy for model-C^{3w} is underestimated by 0.08 eV, a value that can be attributed to the cytoplasmic orientation of Arg82. With its positive excess charge closer to the Schiff base, and further away from the retinal β-ionone ring. The excitation energy of model-C^{3w} is significantly blue-shifted when the Arg82 side-chain extracellularly oriented as in model-A^{3w}.

The active site of the structure prepared from the Asp85Ser mutant[274] with three water molecules (model-B^{3w}) has a rather unusual interaction between the retinal C_ε atom and Asp85. Due to the interaction between retinal and Asp85, the excitation energy of model-B^{3w} is strongly blue-shifted relative to the experimental value. From all structural models investigated here, the experimental excitation energies are best reproduced by model-A^{3w}. Model-A^{3w} also reproduces the [O]-O shift in the excitation energy. However, the excitation energies are not sensitive to large scale conformational changes that occur in regions remote from the active site.

7.7.4 IR Spectral Fingerprints

We used the approximate method SCC-DFTB to compute vibrational frequencies. Using gas-phase models of acetic acid, SCC-DFTB for ν_{COOH} was benchmarked for two extremes scenarios, acetic acid in the gas phase model and a strongly hydrogen bonded cyclic dimer (Table 7.1). We discuss all vibrational frequencies results with respect to these gas phase reference calculations. SCC-DFTB reproduces the monomer C=O stretch frequencies extremely well (Table 7.1), deviating only by 2 cm^{-1} relative to the experimental value[258] of 1780 cm^{-1}. The agreement between experiments and theory is worse for the cyclic dimer, where we found a deviation

of 17 cm^{-1}. These two test cases are the two extremes of a free C=O group, and of a strongly hydrogen bonded acetate, that mark all hydrogen bonding possibilities in the protein. We expect SCC-DFTB to produce increasingly red-shifted vibrational frequencies with increasing hydrogen bonding interactions. This allows us to assess the O and [O] models with respect to the C=O stretch frequencies. The C=O stretch frequencies do not change when salt bridge forms between the Schiff base and Asp85 or Asp212, being sensitive only to direct hydrogen bonding interactions of the protonated aspartate. The main structural determinants of the aspartate stretching frequencies are the hydrogen bonds with active site water molecules and to nearby groups such as Thr89 (Asp85) and Tyr57 (Asp212). For most models, the [O] state vibrational frequencies are in good agreement with the experimental values, whereas the O state frequencies vary significantly, mostly due to the Thr89:Asp85 hydrogen bond.

In the models used in our study, we find that Thr89 hydrogen bonds either to the side-chain (in model-A structures derived from acid-blue bacteriorhodopsin[275]) or to the backbone of Asp85 (in model-B structures derived from the Asp85Ser mutant[274]). Since the presence of the Thr89:Asp85 hydrogen bond red-shifts the C=O frequency by approximately 20 cm^{-1}, the hydrogen bonding of Thr89 (to the carboxyl or to the carbonyl group of Asp85) is clearly distinguished by the vibrational frequency of Asp85. Based on the critical examination of the SCC-DFTB performance for vibrational frequencies, we expect slightly red-shifted frequencies compared to experiment (1752-1756 cm^{-1}). Therefore, computation of vibrational frequencies further supports the model-A^{3w} with the Thr89:Asp85 hydrogen bond. The shift in the ν_{COOH} band for O and [O] states is best reproduced in models with three active site water molecules.

Resonance Raman studies have indicated that the all-*trans* retinal in the O intermediate has a twisted geometry.[272,500] The O-like models used in our study agree with the experimental findings,[272] indicating a larger twist in the retinal geometry in O as compared to bR. The retinal twist occurs during the optimization of model-C^{3w}. To obtain a twist of the retinal chain it is sufficient to start from the bR and change the protonation states to correspond to the O state.

7.8 Conclusion

We modeled the active site of the O state structure using QM/MM simulations. Our investigation of the details of the hydrogen bonded network in the active site indicate that water molecules not present in acid-blue bacteriorhodopsin,[275] Asp85Ser mutant[274] crystal structures, have an important structural role in the O intermediate. The geometric and the electronic (absorption energies) features of the retinal depend on the number of water molecules in the active site. Although the crystal structures of the Asp85Ser mutant[274] and of the acid-blue bacteriorhodopsin[275] indicate only one active site water molecule, the geometry of active site from the crystal structure of acid-blue bacteriorhodopsin[275] is best preserved in computations when three water molecules are present in the active site. Similar to the bR,[21] in O, the Schiff base hydrogen bonds directly to a water molecule. The absence of the water molecule hydrogen bonding to the Schiff base would lead to large blue-shifts in the optical absorption spectra. The Thr89:Asp85 hydrogen bond present in the first half of the photocycle[21,119,157,164,173,206,298] is absent in some of the O state models investigated here. The agreement between the experimental and the calculated C=O stretch frequency computed with the acid-blue bacteriorhodopsin structure in which the Thr89:Asp85 hydrogen bond is present indicates that the Thr89:Asp85 hydrogen bond is indeed a characteristic of the O state. Consistent with the findings of Resonance Raman experiments, all O state models indicate an increased twist of the retinal with respect to the bR.

To summarize, we propose that the O state model derived from the acid-blue bacteriorhodopsin structure[275] with three active site water molecules, a hydrogen bond between Thr89 and Asp85, and twisted retinal is a good representation of the wild-type O state.

Standard QM/MM excitation energy calculations seem to be insufficient for describing the change in absorption energies between the bR and O state since protein polarization effects are very different in magnitude in these two states. The dependency of the effect of polarization on the structural models used indicates that the polarization effects must be carefully taken into account in studies of color tuning in rhodopsins.

According to the reaction path calculations, a PT from Asp85 to Asp212 is thermodynamically feasible considering the rate-limiting energy barrier of $\sim$15 kcal/mol for the O$\rightarrow$bR reaction.[80,501] The putative [O], as an intermediate be-

tween O and bR, shows changes in the spectroscopic properties as reported experimentally.[126,273] The experimental values of the shift in the C=O stretch frequency is $\sim$40 cm^{-1} and the shift in the optical excitation energy is $\sim$0.04 eV between O and [O] states. These changes are well reproduced by our structural models of the O and [O] states modeled from the crystal structure of ref 275 (model-A^{3w}), thereby confirming the picture of a transient O-intermediate, which is characterized by a protonated Asp212 and deprotonated Asp85. These results indicate that indeed, the spectral shift observed in ref 126 is likely due to Asp212 becoming protonated. The calculations support the proposal of ref 126 that Asp212 is an intermediate carrier of the proton during the O$\rightarrow$bR transition.

The mechanism of the complete O$\rightarrow$bR will be the focus of future work. Several key issues must be addressed. For example, the present computations indicate that in O, the Arg82 side-chain is oriented towards the extracellular side. For the recovery of the bR, the side-chain must orient back to its cytoplasmic configuration. The rate-limiting energy barrier, the sequence of events associated with the reorientation of the Arg82 side-chain, and whether or not reorientation of Arg82 is part of the last PT step, are still open questions that will be addressed in the future using the approach presented here.

Chapter 8

Long-range PT during O→bR

Summary

This chapter summarizes results of the preliminary calculations for the long-range PT in bacteriorhodopsin that takes place from Asp85 *via* Asp212 to the PRG. The path calculations with a cytoplasmically oriented Arg82 side chain are described in section 8.3 followed by results from a classical MD simulation in section 8.4. Finally in section 8.5, the results for a path calculation through a chain of hydrogen bonded water molecules are presented.

8.1 Introduction

Less is known about the O→bR transition in comparison to the photocycle transitions that take place in the early half of the bacteriorhodopsin photocycle. The PT during the O→bR transition takes place from Asp85 to the extracellular PRG over ∼12 Å distance (Figure 8.1) and is therefore challenging to investigate. For understanding the mechanism of the PT, it is necessary to investigate several structural and energetic factors that govern the transition.

Kinetic studies have indicated that a large decrease in free energy of ∼15 KJ/mol (i.e., ∼4 kcal/mol) is associated with the recovery of the bR ground state following a strongly downhill O→bR transition.[80,191] The rate of recovery of the bR ground state has also been shown to be coupled with the rate of deprotonation of Asp85 during the O→bR transition.[233] The driving force of this reaction is likely to be the reestablishment of the low pK_a of Asp85 (∼2.5) and the high pK_a of the

PRG (∼9).[502]

FTIR studies of wild-type and several mutant bacteriorhodopsin structures[126,463] suggested that the O→bR transition may involve a transient O state (denoted as [O]) characterized by deprotonated Asp85 and protonated Asp212. Detailed structural and spectroscopical analysis based on the hybrid QM/MM simulations carried out in the current work support these suggestions (Chapter 7). The results suggest that Asp212 is likely to be involved in the pathway of the proton between Asp85 and the PRG.

If [O] state is indeed sampled during the O→bR transition, then the recovery of the bR ground state would take place following the PT from Asp212 to the PRG. Structural and energetic details of the PT from Asp212 to the PRG during [O]→bR remains unclear. It is not clear if the transfer of proton from Asp212 to the PRG could take place through a hydrogen bond network of water molecules as in a Grotthuss mechanism or if the PT could involve participation of Arg82 side chain. However in none of the available O-like structures[274,275] a hydrogen bonded network of water molecules is observed.

The exact role of Arg82 in the O→bR transition also remains unclear. In the available O-like structures,[274,275] Arg82 side chain adopts an extracellular orientation while in the bR ground state it adopts a cytoplasmic orientation (see Figure 7.2c). Therefore, the O→bR transition must involve orientation of Arg82 side chain from extracellular to a cytoplasmic configuration. However, it is not known at what stage this orientation takes place.

The focus of this chapter is to summarize preliminary results of the simulations of PT from Asp85/Asp212 to the PRG that could take place during [O]→bR transition. The simulations indicate that Arg82 is unlikely to be involved in the PT and the excess proton is likely to be transferred through a hydrogen bonded network of water molecules to the PRG.

8.2 Computational Details

Hybrid SCC-DFTB/CHARMM minimum energy path calculations were carried out using same setup as mentioned in chapter 7 PT path calculations were carried starting from two structural models: (i) model-C^{3w} in which Arg82 has a cytoplasmic orientation (see chapter 7), and (ii) model-E^{3w} starting from model-D^{3w}

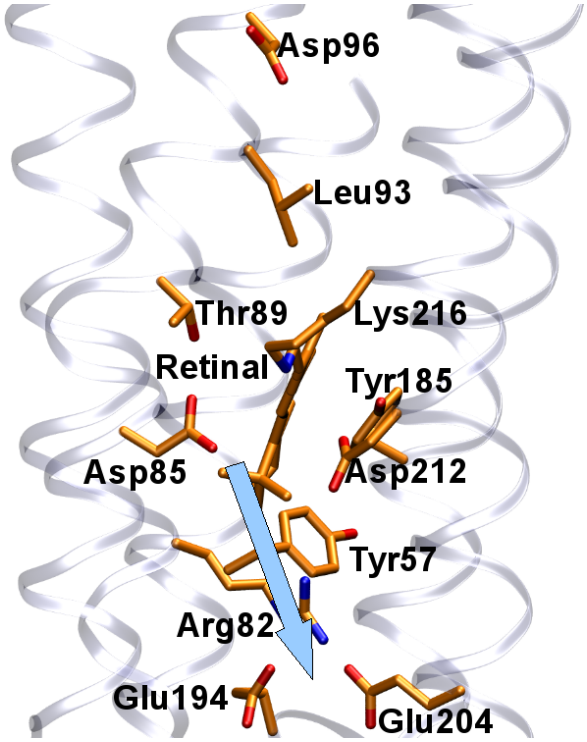


Figure 8.1: The last PT in bacteriorhodopsin takes from Asp85 to PRG over ∼12 Å distance and restores the ground state.

in which Arg82 has an extracellular orientation (see chapter 7) with a hydrogen bonded network of extra water molecules that extend from Asp212 to Glu204.

Asp96 and Asp115 were modeled neutral as indicated by experiments.[126,273] Standard protonation states were used for all remaining titratable protein amino-acids. All buried water molecules present in the crystal structures were included in the calculations. The hydrogen atoms were built using the HBUILD utility of CHARMM package.[357]

For model-C^{3w}, the QM region consisted of Lys216, complete retinal molecule, the side chains of Asp85, Asp212, Arg82, Glu194, and Glu204 and the active site water molecules adding up to 145 atoms. Since QM/MM minimum energy path calculations with 145 QM atoms are very time consuming, for model-E^{3w}, the QM region was restricted to include side chains of Asp85, Asp212, Glu194 and Glu204 and all active site water molecules. Retinal and side chains of Lys216 and Arg82

were treated at MM level.

The boundary between the quantum and classical subsystems was chosen by adding the link atom at the C_α-C_β covalent bond for Asp85, Asp212, Glu194 and Glu204. The QM/MM frontier was treated using DIV linking scheme.[374] To maintain the shape of the protein close to the crystal structure a part of the protein backbone was fixed to the crystal structure coordinates. The remaining part was mobile and consisted of ∼1590 atoms including the quantum mechanical region and a layer of protein groups and surrounding water molecules that can be related to the PT path. The SCC-DFTB/CHARMM energy minimizations were performed to a RMS energy gradient of 10^{-3} kcal/mol.Å.

8.3 PT *via* Arg82 (Path 1)

In model-C^{3w}, the PT path from Asp85 to the PRG through Arg82 side chain. Important conformers during the PT are shown in Figure 8.2. In the reactant (Figure 8.2a), the donor Asp85 and acceptor Glu204 (as a part of the PRG) are well connected through a hydrogen bonded network that involves Asp212, Arg82 and active site water molecules. The pathway involves transient protonation of Asp212 from Asp85 (Figure 8.2b) following which, Asp212 deprotonates to Arg82 completing the PT (Figure 8.2c).

The complete PT can be divided into three concerted pathways (see panels A-C in Figure 8.3). The pathway begins with a transfer of proton from Arg82 to Glu204 *via* water molecule w403 forming neutral Arg82 (panel A in Figure 8.3). A proton is then transferred between nitrogen atoms of Arg82 guanidinium group through w407 (panel B in Figure 8.3). This PT gives rise to the rate-limiting energy barrier of 35.6 kcal/mol (at λ=0.33 in Figure 8.3). The primary donor Asp85 then loses its proton to Asp212 through water molecules w401 and w406 in a concerted manner forming a transient state with anionic Asp85 and neutral Asp212 (Figure 8.2b). The pathway is then completed with PT from Asp212 to neutral Arg82 through water molecule w406 (panel C in Figure 8.3) forming a bR-like product state with anionic Asp85 and protonated PRG (Figure 8.2c).

Considering that the O→bR transition takes place on a ∼ms time-scale, the high rate-limiting barrier of 35.6 kcal/mol suggest that the PT is unlikely to take place in a structure with a cytoplasmic orientation of Arg82 side chain.

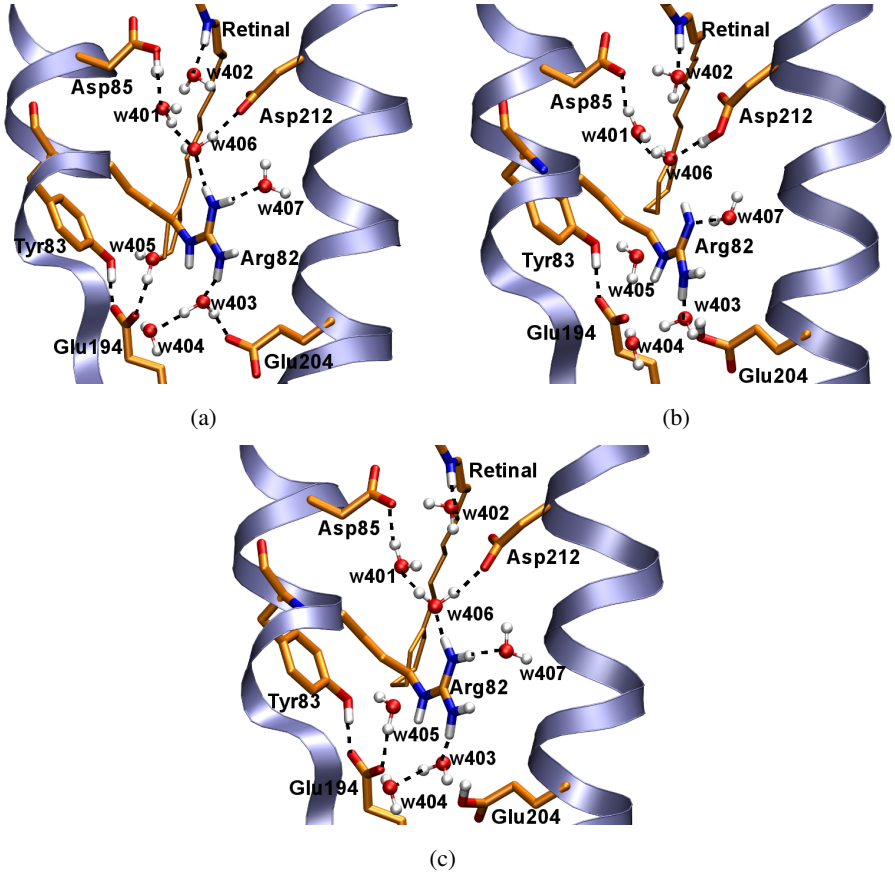


Figure 8.2: Conformation states during PT with model-C^{3w} prepared from ref 21. The water molecules are numbered according to ref 21. (a) Reactant (O-like) state of with protonated Asp85 and anionic Asp212 and Glu204; (b) An intermediate state with anionic Asp85, anionic Asp212, neutral Arg82; (c) Product (bR-like) state with anionic Asp85 and neutral Glu204.

8.4 Classical MD Simulation with Model-C^{3w}

Another possibility for the PT from Asp212 to the PRG can be to transfer a proton through a hydrogen bonded chain of water molecules that would extend from Asp212 to the PRG. However, with cytoplasmic orientation of Arg82 side chain, such a chain of water molecules was not observed in model-C^{3w}. Therefore, additional classical MD simulation with model-C^{3w} was carried out to: (i) explore the structural stability of model-C^{3w} with respect to the orientation of Arg82 side

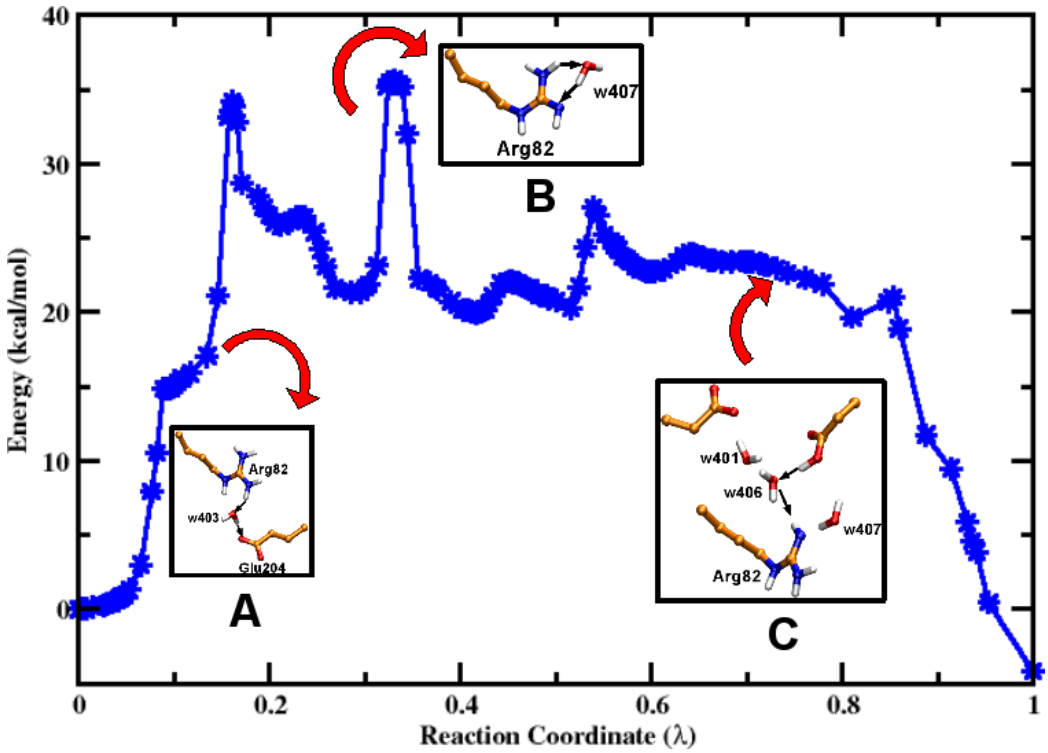
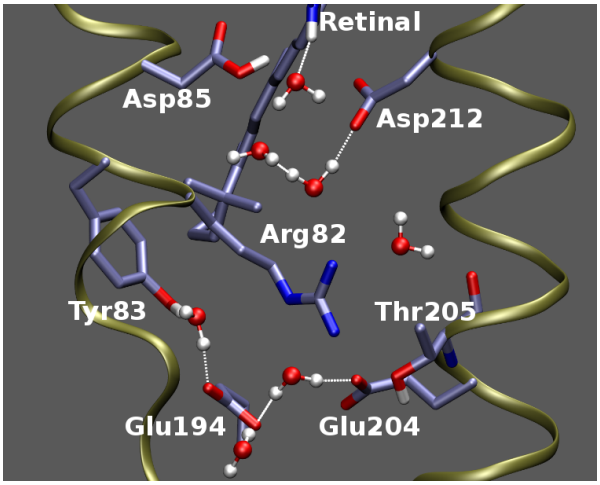


Figure 8.3: Energy profile for the PT from Asp85 to Glu204 (modelled as a part of the PRG). The complete PT consists of mainly three concerted PT pathways: (A) from Arg82 to Glu204 through w403; (B) between nitrogen atoms of Arg82 through w407, and (C) from Asp212 to Arg82.

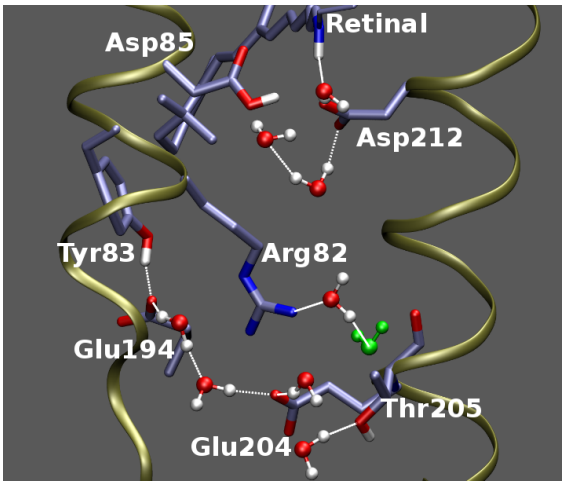
chain, and (ii) investigate if a stable hydrogen bonded network of water molecules can be formed in model-C^{3w}.

A classical MD simulation for 2 ns duration was carried using similar GSBP/MM setup as described in section 7.2.5 with 22 Å GSBP inner region centered around N_ε atom of Arg82. All bonds involving hydrogen for were constrained using SHAKE[493] to allow a 1-fs timestep.

Throughout the simulation, three water molecules in the vicinity of the Schiff base remain in the binding pocket indicating that the water molecules are stable over the sampled 2 ns time-scale. During the first half of the simulation, Arg82 side chain continues to adopt a cytoplasmic orientation (Figure 8.4a) and after ∼450 ps, changes orientation towards the extracellular Glu194/Glu204 protein groups and maintains the extracellular orientation for the rest of the simulation (Figure 8.4b).



(a)



(b)

Figure 8.4: Representative snapshots from classical MD simulation for model-C^{3w} (2 ns trajectory) indicating Arg82 side chain orientation: (a) cytoplasmic; (b) extracellular. During the simulation, an additional bulk water molecule (represented in green) was observed to move into the active site.

Following the extracellular orientation, an additional bulk water molecule moves into the active site and remains in the active site for the rest of the simulation.

Although the classical MD simulation described here was for relatively short duration (2 ns), the simulation does indicate some interesting features: (i) The

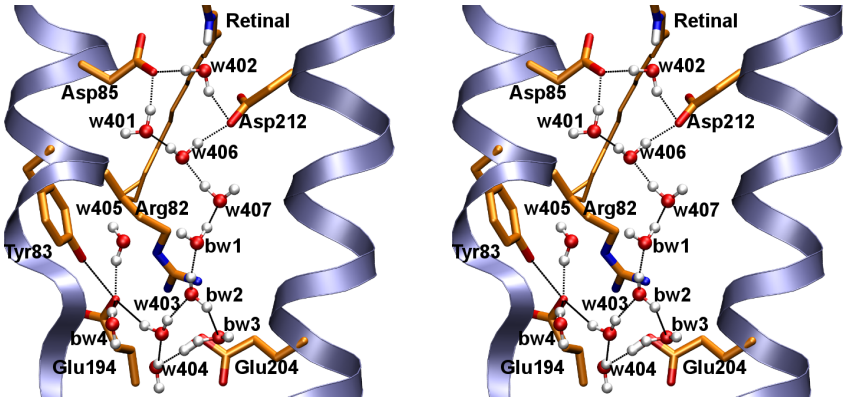
fact that Arg82 favors an extracellular orientation in O-like state of model-C^{3w} indicates that it Arg82 unlikely to change its extracellular orientation back to cytoplasmic orientation (as observed in bR ground state) before the PT is completed during the O→bR transition. (ii) During ∼0.5 ns time-scale over which Arg82 side chain adopts a cytoplasmic orientation, no additional water molecules were observed in the active site. On the other hand, observation of an additional water molecule indicate that it is likely that in the O state additional water molecules can be accommodated in the active site. The water molecules would then form a stable hydrogen bonded network through which a proton can be transferred. The results of this preliminary simulation agree well with the simulation results presented in chapter 7 which indicated formation of a stable chain of water molecules in the O-like conformer.

8.5 PT *via* Water Molecules (Path 2)

To investigate the PT scenario through a hydrogen bonded water chain with extracellular orientation of Arg82, an O-like conformer was prepared starting from model-D^{3w} in which additional water molecules were added (Figure 8.5a). The donor Asp85 is connected to Glu204 through a hydrogen bonded network of water molecules (Figure 8.5a).

The PT in path 2 takes place in a concerted fashion. In the beginning of the pathway, Asp85 donates its proton to nearby w401. Concomitantly, w401 shares a proton with w406 until the excess proton is shared between the water molecules of the active site. The rate-limiting energy barrier of 21.9 kcal/mol (Figure 8.6) thus arises as a result of the shared excess proton between the hydrogen bonded water molecules. Consequently the proton is transferred to Glu204 completing the pathway.

Unlike path 1, the PT in path 2 is endothermic. This is striking since O→bR transition is believed to largely exothermic.[80] The reason for this discrepancy may be due to the hysteresis effect in the calculation of minimum energy pathways using CPR. Further, this last PT is dominated by entropic contributions and hence free energy profile will be a proper measure to understand the PT energetics. Nevertheless, these preliminary calculations give interesting insight into the complex PT path.



(a) (b)

Figure 8.5: Conformation states during PT with model-E^{3w} prepared from ref 21. The water molecules are numbered according to ref 21. Four additional bulk water molecules are denoted as bw1 to bw4. (a) Reactant (O-like) state of with protonated Asp85 and anionic Asp212 and Glu204; (b) Product (bR-like) state with anionic Asp85 and neutral Glu204.

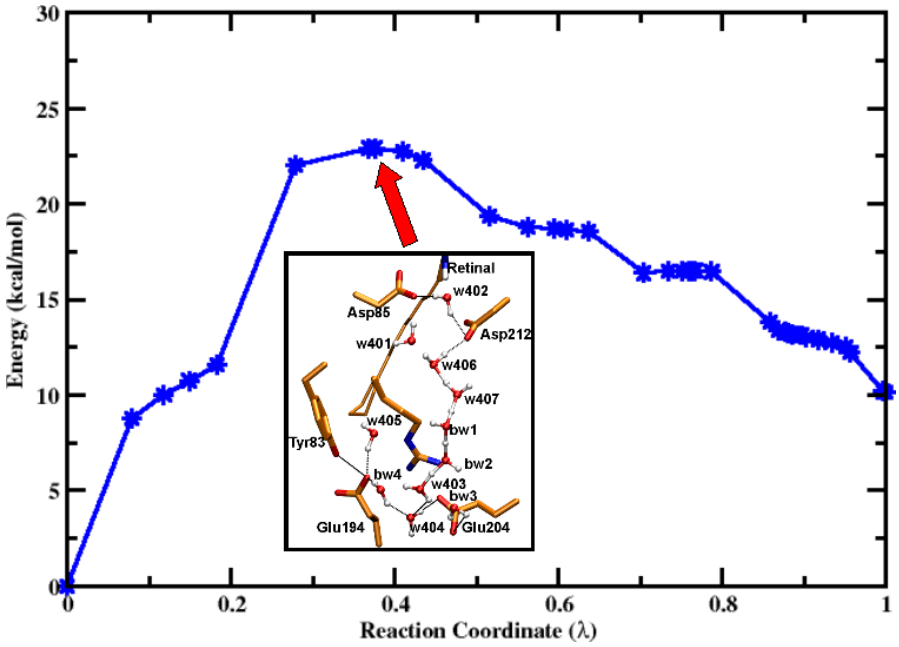


Figure 8.6: Energy profile for the PT from Asp85 to Glu204 (modelled as a part of the PRG) through a chain of hydrogen bonded water molecules. The rate-limiting barrier of 21.9 kcal/mol (at λ=0.37) is due to a shared proton between the water molecules.

8.6 Discussion

Although the results presented for the long-range PT are preliminary, the calculations does shed some light on the complex PT scenario during the O→bR transition. The rate-limiting energy barrier of ∼36 kcal/mol in path 1 indicates that the PT during the O→bR transition is less likely to take place with a cytoplasmically oriented Arg82 side chain. Further, a short classical MD simulation indicates that Arg82 side chain is more likely to adopt to an extracellular orientation in O state prior to the PT which agrees well with the available O-like structures.[274,275]

Thus with extracellular orientation of Arg82 side chain, the proton may be transferred through a chain of water molecules that extend from Asp212 to Glu204. Although, in none of the available O-like structures,[274,275] such a hydrogen bonded water chain is observed, classical MD simulations (see section 7.3.3) suggest that a chain of water molecules can be formed in O conformer over 5 ns time-scale and thus support this possibility of PT through water molecules. The path 2 simulations also support such a possibility.

The O→bR transition takes place on a milli-second time-scale and hence the entropic and solvent effects on the reaction are important. Consequently, the free energy simulations are warranted for proper treatment of these effects and thus could provide a better understanding of the mechanism of PT. This will be the focus of future work.

Chapter 9

Epilogue

Energy is one of the most fundamental quantity in nature and it is almost impossible to imagine the universe in absence of it. All biological as well as non-biological phenomena utilize energy in certain ways governed by the principle of conservation of energy. Mankind has spent substantial amount of *energy* to conceive better and better understanding of forms of energy and processes associated with energy transformations.

Of particular interest are biological processes which consume energy to carry out several biochemical reactions that are crucial for sustaining life. An archaeon *Halobacterium salinarum* employes a transmembrane protein, bacteriorhodopsin to generate proton-motive gradient across the cell membrane which is subsequently used for ATP synthesis.

The aim of the current work was to understand the mechanism of the proton pumping in bacteriorhodopsin by employing computer simulation techniques. Multi-length-scale simulations were carried out to investigate mainly two important issues of the light-activated photocycle. The findings from the current work agree well with available experimental results.

Investigations for the identity of the proton release group (PRG) were carried out in this work and the obtained results support the earlier proposed model of the PRG to be a shared proton between Glu194 and Glu204 residues. The simulations indicate that an excess proton on the protonated water cluster is highly unstable and moves within a few picoseconds to become delocalized between Glu194/Glu204 residues. Such a delocalized proton between Glu194/Glu204 residues gives rise to a continuum band in the 1800-2000 cm^{-1} region in agreement with observed in FT-

IR experiments. Another important consequence of the strong hydrogen bonding interaction between Glu194/Glu204 residues is the strong red-shift of the ν_{COOH} band of Glu194/Glu204 to below 1700 cm^{-1} region.

The simulations also indicate that the active site geometry with a delocalized proton between Glu194/Glu204 residues agrees well with the available crystal structural models. On the other hand, the simulations in which an excess proton was stabilized on the protonated water cluster led largely perturbed active site geometry and thus support the observation that the protonated water cluster is unlikely to function as the PRG in the bacteriorhodopsin photocycle.

For a complete understanding of the long-range PT during O→bR transition, it is essential to investigate the structural features of the O state as well as to assess whether O→bR transition could involve any transient state. Detailed structural and spectroscopical analysis that was carried out in this work suggest that in the O state, the retinal is twisted in O and three water molecules are likely to be present in the retinal active site.

The simulations suggest that a transient O state (denoted here as [O]) could be sampled during O→bR transition and it is characterized by anionic Asp85 and neutral Asp212. The calculated shifts between O and [O] conformers with respect to the vertical excitation energy (0.07 eV) and ν_{COOH} band (38 cm^{-1}) agree well with the experimental shifts of 0.04 eV and 42 cm^{-1} respectively.

In addition to the PT from Asp85 to the PRG, the last photocycle step in bacteriorhodopsin is associated with re-orientation of Arg82 from an extracellular to a cytoplasmic configuration. Preliminary simulations were carried out to understand the effect of Arg82 orientation and its involvement on the long-range PT. The results indicate that Arg82 is less likely to be involved in the PT pathway from Asp212 to the PRG. Thus, the PT to the PRG is likely to take place through a chain of hydrogen bonded water molecules that can be formed with an extracellular orientation of Arg82.

The results obtained in this work thus emphasize the importance of proper computational setup as well as the necessity of an understanding of the available spectral and structural features of bacteriorhodopsin to elucidate the mechanism of PT in bacteriorhodopsin. The findings also illustrate the strength of computer simulations to reveal the origin of observered features to the atomistic details. To summarize, the present work is an attempt to elevate the confidence in the advanced

simulation techniques as a tool to explore complex biological processes.

9.1 Outlook

It is generally believed that the expansion of the universe began about 14 billion years ago and it still continues today. Over the years, the immense desire of mankind to understand the properties and phenomena from the universe has led to a substantially large number of discoveries and inventions. Nevertheless if one would pause for a moment to ponder how close is the present state of knowledge from the ultimate understanding of everything, it would be immediately apparent that its still a long way to go. Can the reason for this large and often disturbing difference be attributed to the inherent expanding nature of the universe? If this is true, then would it mean that, more and more expansion of the universe will lead to more and more unsolved problems, mysteries?

Irrespective of whether above argument is sensible or not, above mentioned knowledge difference is apparent on all research fronts including the investigation of the mechanism of bacteriorhodopsin. An attempt is made in this section to summarize the open issues with respect to the last photocycle step that takes place during O→bR transition which will be the focus of future work.

The origin of the driving force for Asp85 deprotonation during O→bR transition remains unclear. One possibility to understand the origin may be to carry out pK_a simulations. The simulations will be aimed at understanding the influence of protein groups on pK_a of donor Asp85 and acceptor PRG.

Preliminary calculations on the long-range PT (chapter 8) presented in this work indicate that Arg82 is likely to change its orientation following the PT. To investigate the energetic and structural changes associated with this movement, free energy calculations will be carried out. The stability of the observed chain of hydrogen bonded water molecules in the extracellular active site of O will be investigated in detail using grand canonical Monte Carlo simulations.

The last photocycle step is one of the slowest transition in the bacteriorhodopsin photocycle and it is believed to be associated with significant changes in entropy and consequently free energy of the system. To understand these entropic and solvent effects on the long-range PT path from Asp85 to the PRG, free energy umbrella sampling simulations will be carried out.

List of Abbreviations and Programs

2D	Two dimensional
3D	Three dimensional
bR	Bacteriorhodopsin ground state
CP	Cytoplasmic side
EC	Extracellular side
fs	femto second
FTIR	Fourier transform infrared spectroscopy
GSBP	Generalized solvent boundary potential
MD	Molecular dynamics
MM	Molecular mechanics
ms	milli second
NMR	Nuclear magnetic resonance spectroscopy
ns	nano second
PDB	Protein data bank
ps	pico second
PT	Proton transfer
PRG	Proton release group
QM	Quantum mechanics
SORCI	Spectroscopically oriented configuration interaction
μs	micro second

List of Programs

- **Protein Data Bank:** All experimentally resolved crystal structures of various bacteriorhodopsin intermediates are available from the Protein Data Bank (PDB) accessible on the internet at http://www.pdb.org.
- **CHARMM:** All simulations in this work were carried out using CHARMM simulation package.
- **Gaussian:** The IR benchmark calculations were carried out using Gaussian03 package (http://www.gaussian.org)
- **Pymol and VMD:** For visualisation of structures and for preparation of pictures, Pymol and VMD softwares were used. Pymol and VMD packages are available from the website at http://pymol.sourceforge.net and http://www.ks.uiuc.edu/Research/vmd respectively.
- **Xmgrace and Origin:** All the graphs in the current work were prepared using Xmgrace and Origin packages. Xmgrace is available at http://plasmagate.weizmann.ac.il/Grace and Origin is available http://www.originlab.com
- **$\mathbf{L^AT_EX}$:** The thesis was typesetted using L^AT_EX (available at http://www.latex-project.org) in the Kile-1.8 integrated environment for L^AT_EX. The Kile package is available at http://kile.sourceforge.net

Acknowledgments

Since the beginning of my doctoral work, I have always felt that one of the most difficult part of the doctoral work is to properly acknowledge everyone associated with the work. This is perhaps because the human life evolves through complex and often overlapping relationships between individuals on personal and professional level and which are often difficult to separate. Although the laws of nature usually work well for understanding and predicting phenomena around us, they seem so inadequate when it comes to the differentiation of diverse interactions of individuals with each other. Nevertheless, I shall attempt to the best of my ability to acknowledge people who made this doctoral work possible. Nevertheless there is a good probability that many friends and colleagues may be left out unintentionally. I would like to express my sincere thanks to all those who have supported and helped me during my work.

I would like to begin from the beginning of the universe. Irrespective of the origin of the universe, I would like to thank to that unclear and often disputed entity or quantity that made existence of vivid and exotic life on earth possible. If the interactions of earth with other celestial bodies had been little different, perhaps neither we nor other living organisms would have flourished.

Coming back to earth, this doctoral work will be complete only with my sincere gratitude to my guide Prof. Dr. Marcus Elstner. I wish to thank him for accepting and supporting me to work on the interesting and challenging area of bacteriorhodopsin simulations. I shall always admire the clarity and simplicity with which he is able to discuss diverse topics ranging from fundamentals of quantum mechanics to complex biochemical reactions. Although a certain amount of uncertainty remains associated with his physical whereabouts, I thank him for the way in which he made me feel his presence as a colleague, mentor and friend.

Marcus not only guided me to tackle numerous methodological and personal problems, but also helped me to conceive the accuracies and limitations of simulation techniques.

I thank the Deutsche Forschungsgemeinschaft for the research grant of the "Forschergruppe 490 Molecular Mechanisms of Retinal Protein Action" for financial support. I also thank Prof. Dr. S. Suhai from DKFZ, Heidelberg for allowing me to work in his group as a visiting scientist during the early part of my doctoral work. I acknowledge and thank people responsible for properly maintaining the computational resources from DKFZ, IWR Heidelberg, University of Paderborn, and TU-Braunschweig. Without these resources this doctoral work would not have been possible.

I would like to thank my teachers especially, Dr. Dabhade, Dr. R. S. Joshi from Fergusson college Pune, and Prof. S. H. Patil, Prof. U. Yajnik and Prof. P. Ramadevi from Indian Institute of Technology, Bombay who taught me numerous courses in basic sciences. Special thanks to Prof. Liisa Holm at University of Helsinki for introducing me to the field of bioinformatics and biomolecular simulations.

I am indebted to my colleagues Dr. Paul Strodel and Dr. Ana-Nicoleta Bondar who not only helped getting started with the simulations, but also for aspiring scientific discussions and collaborations. I thank Dr. Michael Hoffmann and Dr. Peter König for interesting discussions especially with respect to charmm simulations. I would like to thank Marius Wanko from BCCMS Bremen for the ongoing collaboration on retinal proteins and for fruitful discussions on the excitation energy calculations.

My sincere thanks for Prof. Dr. Qiang Cui and Dr. Nilanjan Ghosh from University of Wisconsin, Madison for the fruitful and ongoing collaboration on the proton release group. Special thanks for Dr. Nilanjan Ghosh for his help in setting up the GSBP/CHARMM simulations.

On the administrative front, I thank Anke Retzmann from DKFZ, Simone Lange from University of Paderborn and Brigitte Schäfer from TU-Braunschweig for taking care of all the involved paperwork with respect to work contract, residence permit and travel grants. I thank Illona Schweda from Promotionsstelle, and co-workers from the International office of TU-Braunschweig for taking care of the documents necessary for the submission of my dissertation.

Special thanks to Dr. Christian Bleiholder from DKFZ for interesting discussions during lunch-time in the Heidelberg university mensa. I thank my colleagues from theoretical chemistry group from TU-Braunschweig for creating lively and interesting work environment and for discussions on scientific as well as non-scientific topics. I thank Jan Frähmcke, Dr. Tomáš Kubař and Michael Gaus for their critical suggestions which helped me in improving the dissertation.

I owe a sincere gratitude to my colleague and friend Jan Frähmcke for helping me on all fronts during my stay in Braunschweig right from finding an apartment to stimulating scientific discussions. The ongoing collaboration with him on simulation of retinal proteins has been very fruitful and inspiring.

I would also like to thank two of my closest friends Dr. Abhinav Verma and Dr. Harshad Joshi. Long discussions with Abhinav and Harshad covered virtually all topics and I thank them for all the great times we shared during their stay in germany and for their help and support on all fronts of my scientific and personal life. I also wish to thank my other Indian friends Sravan, Amol, Prajakta, Asmita and Abhijit for their support through happy and difficult times during my stay in Braunschweig.

Finally on the family front, I would express my deepest and sincere gratitude to my beautiful and loving wife Anjali for her constant care, support and encouragement. The work would not have been possible without persistent support and love of my parents and sister. I thank my family members for teaching me lessons of life, feeding me and motivating me to do good and useful work. I will always remain indebted to Anjali, my caring sister Rajashree, mom and dad.

I dedicate this thesis to my family and to my teachers...

Bibliography

[1] Lohmann, K. *Naturwissenschaften*, 1929, 17:624–625.

[2] Fiske, C. H. and Subbarow, Y. *Science*, 1929, 70:381–382.

[3] Mitchell, P. *Nature*, 1961, 191:144–148.

[4] Schaefer, G.; Engelhard, M. and Mueller, V. *Microbiol. Mol. Biol. Rev.*, 1999, 63:570–620.

[5] Woese, C. R. and Fox, G. E. *Proc. Natl. Acad. Sci. USA*, 1977, 74:5088–5090.

[6] Spudich, J. L.; Yang, C.; Jung, K. and Spudich, E. N. *Annu. Rev. Cell Dev. Biol.*, 2000, 16:365–392.

[7] Ng, W. V.; Kennedy, S. P.; Mahairas, G. G.; Berquist, B.; Pana, M; Shukla, H. D.; Lasky, S. R.; Baliga, N. S.; Thorsson, V.; Sbrogna, J.; Swarzell, S.; Weir, D.; Hall, J.; Dahl, T. A.; Welti, R.; Goo, Y. A.; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M. J.; Hough, D. W.; Maddocks, D. G.; Jablonski, P. E.; Krebs, M. P.; Angevine, C. M.; Dale, H.; Isenbarger, T. A.; Peck, R. F.; Pohlschroder, M.; Spudich, J. L.; Jung, K.; Alam, M.; Freitas, T.; Hou, S.; Daniels, C. J.; Dennis, P. P.; Omer, A. D.; Ebhardt, H.; Lowe, T. M.; Liang, P.; Riley, M.; Hood, L. and DasSarma, S. *Proc. Natl. Acad. Sci. USA*, 2000, 97:12176–12181.

[8] Hoff, W. D.; Jung, K. and Spudich, J. L. *Annu. Rev. Biophys. Biomol. Struct.*, 1997, 26:223–258.

[9] Spudich, J. L. *Mol Microbiol.*, 1998, 28:1051–1058.

[10] Stoeckenius W. and Rowen R. *J. Cell Biol.*, 1967, 34:365–393.

[11] Stoeckenius, W. and Kunau, W. H. *J. Cell Biol.*, 1968, 38:337–357.

[12] Oesterhelt, D. and Stoeckenius, W. *Nature New Biol.*, 1971, 233:149–152.

[13] Blaurock, A. and Stoekenius, W. *Nature New Biol.*, 1971, 233:152–155.

[14] Oesterhelt, D. and Stoeckenius, W. *Proc. Nat. Acad. Sci. USA*, 1973, 70:2853–2857.

[15] Oesterhelt, D. and Krippahl, G. *FEBS Lett.*, 1973, 36:72–76.

[16] Danon, A. and Stoeckenius, W. *Proc. Nat. Acad. Sci. USA*, 1974, 71:1234–1238.

[17] Racker, E. and Stoeckenius, W. *J. Biol. Chem.*, 1974, 249:662–663.

[18] Hartmann, R. and Oesterhelt, D. *Eur. J. Biochem.*, 1977, 77:325–335.

[19] Müller, D. J.; Heymann, J. B.; Oesterhelt, F.; Möller, C.; Gaub, H.; Büldt, G. and Engel, A. *Biochim. Biophys. Acta*, 2000, 1460:27–38.

[20] Heyes, C. D. and El-Sayed, M. A. *J. Phys. Chem. B*, 2003, 107:12045–12053.

[21] Luecke, H.; Schobert, B.; Richter, H. T.; Cartailler, J. P. and Lanyi, J. K. *J. Mol. Biol.*, 1999, 291:899–911.

[22] Henderson, R. *J. Mol. Biol.*, 1975, 93:123–138.

[23] Unwin, P. N. T. and Henderson, R. *J. Mol. Biol.*, 1975, 94:425–440.

[24] Henderson, R. and Unwin, P. N. T. *Nature*, 1975, 257:28–31.

[25] Ovchinnikov, Yu. A.; Abdulaev, N. G.; Feigina, M. Yu.; Kiselev, A. V. and Lobanov, N. A. *FEBS Lett.*, 1977, 84:1–4.

[26] Ovchinnikov, Yu. A.; Abdulaev, N. G.; Feigina, M. Yu.; Kiselev, A. V. and Lobanov, N. A. *FEBS Lett.*, 1979, 100:219–224.

[27] Khorana, H. G.; Gerber, G. E.; Herlihy, W. C.; Gray, C. P.; Anderegg, R. J.; Nihel, K. and Biemann, K. *Proc. Natl. Acad. Sci. USA*, 1979, 76:5046–5050.

[28] Engelman, D. M; Henderson, R.; McLachlan, A. D. and Wallace, B. A. *Proc. Natl. Acad. Sci. USA*, 1980, 77:2023–2027.

[29] Trewhella, J.; Anderson, S.; Fox, R.; Gogol, E.; Khan, S. and Engelman, D. *Biophy. J.*, 1983, 42:233–241.

[30] Kates, M. *Techniques of Lipidology: Isolation, Analysis and Identification of Lipids*. 1986, Elsevier Press, Amsterdam-New York.

[31] Heyes, C. D. and El-Sayed, M. A. *J. Biol. Chem.*, 2002, 277:29437–29443.

[32] Weik, M.; Patzelt, H.; Zaccai, G. and Oesterhelt, D. *Mol. Cell*, 1998, 1:411–419.

[33] Joshi, M. K.; Dracheva, S.; Mukhopadhyay, A. K.; Bose, S. and Hendler, R. W. *Biochemistry*, 1998, 37:14463–14470.

[34] Hendler, R. W. and Dracheva, S. *Biochemistry (Moscow)*, 2001, 66:1311–1314.

[35] Cartailler, J. P. and Luecke, H. *Annu. Rev. Biophys. Biomol. Struct.*, 2003, 32:285–310.

[36] Corcelli, A.; Colella, M.; Mascolo, G.; Fanizzi, F. P and Kates, M. *Biochemistry*, 2000, 39:3318 – 3326.

[37] Dracheva, S.; Bose, S. and Hendler, R.W. *FEBS Lett.*, 1996, 382:209–212.

[38] Corcelli, A.; Lattanzio, V. M. T.; Mascolo, G.; Papadia, P. and Fanizzi, F. *J. Lipid Res.*, 2002, 43:132–140.

[39] Krebs, M. P. and Isenbarger, T. A. *Biochim. Biophys. Acta*, 2000, 1460:15–26.

[40] Lopez, F.; Lobasso, S.; Colella, M.; Agostiano, A. and Corcelli, A. *Photochem. Photobiol.*, 1999, 69:599–604.

[41] Mukhopadhay, A. K.; Bose, S. and Hendler, R. W. *Biochemistry*, 1994, 33:10889–10895.

[42] Fitter, J.; Verclas, S. A. W.; Lechner, R. E.; Seelert, H. and Dencher, N. A. *FEBS Lett.*, 1998, 433:321–325.

[43] Dencher, N. A.; Sass, H. J. and Büldt, G. *Biochim. Biophys. Acta*, 2000, 1460:192–203.

[44] Falk, K. E.; Karlsson, K. A. and Samuelsson, B. E. *Chem. Phys. Llipids*, 1980, 27:9–21.

[45] Lind, C.; Hojeberg, B. and Khorana, H. G. *J. Biol. Chem*, 1981, 256:8298–8305.

[46] Teissie, J.; Prats, M.; LeMassu, A.; Stewart, L. C. and Kates, M. *Biochemistry*, 1990, 29:59–65.

[47] Mowery, P. C.; Lozier, R. H.; Chae, Q.; Tseng, Y. W.; Taylor, M. and Stoeckenius, W. *Biochemistry*, 1979, 18:4100 – 4107.

[48] Chang, C. H.; Chen, J. G.; Govindjee, R. and Ebrey, T. *Proc. Natl. Acad. Sci. USA*, 1985, 82:396–400.

[49] Kimura, Y.; Ikegami, A. and Stoeckenius, W. *Photochem. Photobiol.*, 1984, 40:641–646.

[50] Ariki, M. and Lanyi, J. K. *J. Biol. Chem.*, 1986, 261:8167–8174.

[51] Jonas, R.; Koutalos, Y. and Ebrey, T.G. *Photochem. Photobiol.*, 1990, 52:1163–1177.

[52] El-Sayed, M. A.; Yang, D.; Yoo, S. K. and Zhang, N. *Isr. J. Chem.*, 1995, 35:465–474.

[53] Tan, E. H. L.; Govender, D. S. K. and Birge, R. R. *J. Am. Chem. Soc.*, 1996, 118:2752–2753.

[54] Tuzi, S.; Yamaguchi, S.; Tanio, M.; Konishi, H.; Inoue, S.; Naito, A.; Needleman, R.; Lanyi, J. K. and Saitô, H. *Biophys. J.*, 1999, 76:1523–1531.

[55] Sanz, C.; Márquez, M.; Perálvarez, A.; Elouatik, S.; Sepulcre, F.; Querol, E.; Lazarova, T. and Padrós, E. *J. Biol. Chem.*, 2001, 276:40788 – 40794.

[56] Eliash, T.; Weiner, L.; Ottolenghi, M. and Sheves, M. *Biophy. J.*, 2001, 81:1155–1162.

[57] Sepculcre, F.; Cordomí, A.; Proietti, M. G.; Perez, J. J.; García, J.; Querol, E. and Padrós, E. *Proteins*, 2007, 67:360–374.

[58] Szundi, I. and Stoeckenius, W. *Proc. Natl. Acad. Sci. USA*, 1987, 84:3681–3684.

[59] Szundi, I. and Stoeckenius, W. *Biophy. J.*, 1988, 54:227–232.

[60] Szundi, I. and Stoeckenius, W. *Biophys. J.*, 1989, 56:369–383.

[61] Váró, G.; Brown, L. S.; Needleman, R. and Lanyi, J. K. *Biophy. J.*, 1999, 76:3219–3226.

[62] Lewis, A.; Spoonhower, J.; Bogomolni, R. A.; Lozier, R. H. and Stoeckenius, W. *Proc. Natl. Acad. Sci. USA*, 1974, 71:4462–4466.

[63] Schrekenbach, T.; Walckhoff, B. and Oesterhelt, D. *Eur. J. Biochem.*, 1977, 76:499–511.

[64] Walker, J. E.; Carne, A. F. and Schmitt, H. W. *Nature (London)*, 1979, 278:653–654.

[65] Bayley, H.; Huang, K. S.; Radhakrishnan, R.; Ross, A. H.; Takagaki, Y. and Khorana, H. G. *Proc. Natl. Acad. Sci. USA*, 1981, 78:2225–2229.

[66] Lemke, H. D. and Oesterhelt, D. *FEBS Lett.*, 1981, 128:255–260.

[67] Fahmy, K.; Siebert, F.; Grossjean, M. F. and Tavan, P. *J. Mol. Struct.*, 1989, 214:257–288.

[68] Huang, J. Y. and Lewis, A. *Biophy. J.*, 1989, 55:835–842.

[69] Smith, S. O.; Myers, A. B.; Pardoen, J. A.; Winkel, C.; Mulder, P. P. J.; Lugtenburg, J. and Mathies, R. *Proc. Natl. Acad. Sci. USA*, 1984, 81:2055–2059.

[70] Farrar, M. R.; Lakshmi, K. V.; Smith, S. O.; Brown, R. S.; Raap, J.; Lugtenburg, J.; Griffin, R. G. and Herzfeld, J. *Biophys. J.*, 1993, 65:310–315.

[71] Lozier, R. H.; Niederberger, W.; Ottolenghi, M.; Sivorinowski, G. and Stoeckenius, W. *Energetics and Structure of Halophilic Microorganisms*, chapter On the photocycle of light- and dark-adapted bacteriorhodopsin, pages 129–141. 1979, Elsevier/North-Holland Biomedical Press.

[72] Sperling, W.; Rafferty, C. N.; Kohl, K. D. and Dencher, N. A. *FEBS Lett.*, 1979, 97:129–132.

[73] Birge, R. R. *Biochim. Biophys. Acta.*, 1990, 1016:293–327.

[74] Humphrey, W.; Dalke, A. and Schulten, K. *J. Molec. Graphics*, 1996, 14:33–38.

[75] Rothschild, K. J.; Gray, D.; Mogi, T.; Marti, T.; Braiman, M. S.; Stern, L. J. and Khorana, H. G. *Biochemistry*, 1989, 28:7052 – 7059.

[76] Belrhali, H.; Nollert, P.; Royant, A.; Menzel, C.; Rosenbusch, J. P.; Landau, E. M. and Pebay-Peyroula, E. *Structure*, 1999, 7:909–917.

[77] Zimani, L.; Váro, G.; Chang, M.; Ni, B.; Needleman, R. and Lanyi J.K. *Biochemistry*, 1992, 31:8535–8543.

[78] Brack, T. L. and Atkinson, G. H. *J. Mol. Struct.*, 1989, 214:289–303.

[79] Kouyama, T.; Bogomolni, R. A. and Stoeckenius, W. *Biophy. J.*, 1985, 48:201–208.

[80] Ludmann, K.; Gergely, C. and Váró, G. *Biophy. J.*, 1998, 75:3110–3119.

[81] Mathies, R. A.; Lin, S. W.; Ames, J. B. and Pollard, W. T. *Annu. Rev. Biophys. Biophys. Chem.*, 1991, 20:491–518.

[82] Lanyi, J. K. *Biochim. Biophys. Acta*, 1993, 1183:241–261.

[83] Peteanu, L. A.; Schoenlein, R. W.; Wang, Q.; Mathies, R. A. and Shank, C. V. *Proc. Natl. Acad. Sci. USA*, 1993, 90:11762–11766.

[84] Schoenlein, R. W.; Peteanu, L. A.; Wang, Q.; Mathies, R. A. and Shank, C. V. *J. Phys. Chem.*, 1993, 97:12087–12092.

[85] Mathies, R. A.; Brito Cruz C. H.; Pollard, W. T. and Shank, C. V. *Science*, 1988, 240:777–779.

[86] Hasson, K. C.; Gai, F. C. and Anfinrud, P. A. *Proc. Natl. Acad. Sci. USA*, 1996, 93:15124–15129.

[87] Haran, G.; Wynne, K.; Xie, A.; He, Q.; Chance, M. and Hochstrasser, R.M. *Chem. Phys. Lett.*, 1996, 261:389–395.

[88] Gai, F.; Hasson, K. C.; McDonald, J. C. and Anfinrud, P. A. *Science*, 1998, 279:1886–1891.

[89] Du, M. and Fleming, G. R. *Biophys. Chem.*, 1993, 48:101–111.

[90] Song, L. and El-Sayed, M. A. *J. Am. Chem. Soc.*, 1998, 120:8889–8890.

[91] Atkinson, G. H.; Brack, T. L.; Blanchard, D. and Rumbles, G. *Chem. Phys.*, 1989, 131:1–15.

[92] van den Berg, R.; Jang, D. J.; Bitting, H. C. and El-Sayed, M. A. *Biophy. J.*, 1990, 58:1–15.

[93] Doig, S. J.; Reid, P. J. and Mathies, R. A. *J. Phys. Chem.*, 1991, 95:6372–6379.

[94] Diller, R.; Maiti, S.; Walker, G. C.; Cowen, B. R.; Pippenger, R.; Bogomolni R. A. and Hochstrasser, R. M. *Chem. Phys. Lett.*, 1995, 241:109–115.

[95] Zhong, Q.; Ruhman, S.; Ottolenghi, M.; Sheves, M.; Friedman, N.; Atkinson, G. H. and Delaney, J. K. *J. Am. Chem. Soc.*, 1996, 118:12828.

[96] Atkinson, G. H.; Ujj, L. and Zhou, Y. *J. Phys. Chem. A*, 2000, 104:4130–4139.

[97] Nuss, M. C.; Zinth, W.; Kaiser, W.; Kölling, E. and Oesterhelt, D. *Chem. Phys. Lett.*, 1985, 117:1–8.

[98] Polland, H.; Franz, M. A.; Zinth, W.; Kaiser, W.; Köling, E. and Oesterhelt, D. *Biophys. J.*, 1986, 49:651–662.

[99] Sharkov, A.; Pakulev, A.; Chekalin, S. and Matveetz, Y. *Biochim. Biophys. Acta*, 1985, 808:94–102.

[100] Terentis, A. C.; Ujj, L.; Abramczyk, H. and Atkinson, G. H. *Chem. Phys.*, 2005, 313:51–62.

[101] Goldschmidt, C. R.; Kalisky, O.; Rosenfeld, T. and Ottolenghi, M. *Biophy. J.*, 1977, 17:179–183.

[102] Scheider, G.; Diller, R. and Stockburger, M. *Chem. Phys.*, 1989, 131:17–29.

[103] Tittor, J. and Oesterhelt, D. *FEBS Lett.*, 1990, 263:269–273.

[104] Govindjee, R.; Balashov, S. P. and Ebrey, T. G. *Biophys. J.*, 1990, 58:597608.

[105] Rohr, M.; Gaertner, W.; Schweitzer, G.; Holzwarth, A. R. and Braslavsky, S. E. *J. Phys. Chem.*, 1992, 96:6055–6061.

[106] Logunov, S. L.; El-Sayed, M. A.; Song, L. and Lanyi, J. K. *J. Phys. Chem.*, 1996, 100:2391–2398.

[107] Birge, R. R.; Cooper, T. M.; Lawrence, A. F.; Masthay, M. B.; Zhang, C. F. and Zidovetzki, R. *J. Am. Chem. Soc.*, 1991, 113:4327–4328.

[108] Bondar, A.-N.; Fischer, S.; Suhai, S. and Smith, J. C. *J. Phys. Chem. B*, 2005, 109:14786–14788.

[109] Lozier, R. H.; Bogomolni, R. A. and Stoeckenius, W. *Biophy. J.*, 1975, 15:955–962.

[110] Kalisky, O. and Ottolenghi, M. *Photochem. Photobiol.*, 1982, 35:109–115.

[111] Nagle, J. F.; Parodi, L. A. and Lozier, R. H. *Biophy. J.*, 1982, 38:161–174.

[112] Stoeckenius, W. and Lozier, R. H. *J. Supramol. Struct.*, 1974, 2:769–774.

[113] Milder, S. J. and Kliger, D. S. *Biophy. J.*, 1988, 53:465–468.

[114] Shichida, Y.; Matuoka, S.; Hidaka, Y. and Yoshizawa, T. *Biochim. Biophys. Acta.*, 1983, 723:240–246.

[115] Braiman, M. S. and Mathies, R. A. *Proc. Natl. Acad. Sci. USA*, 1982, 79:403–407.

[116] Aton, B.; Doukas, A. G.; Callender, R. H.; Becher, B. and Ebrey, T. G. *Biochemistry*, 1977, 16:2995–2999.

[117] Ames, J. B.; Fodor, S. P. A.; Gebhard, R.; Raap, J.; van den Berg, E. M. M.; Lugtenburg, J. and Mathies, R. A. *Biochemistry*, 1989, 28:3681–3687.

[118] Gerwert, K.; Souvignier, G. and Hess, B. *Proc. Natl. Acad. Sci. USA*, 1990, 87:9774–9778.

[119] Bondar, A.-N.; Elstner, M.; Suhái, S.; Smith, J. C. and Fischer, S. *Structure*, 2004, 12:1–20.

[120] Kandori, H. *Biochim. Biophys. Acta.*, 2004, 1658:72–79.

[121] Lanyi, J. K. *Annu. Rev. Physiol.*, 2004, 66:665–688.

[122] Hessling, B.; Herbst, J.; Rammelsberg, R. and Gerwert, K. *Biophys. J.*, 1997, 73:2071–2080.

[123] Subramaniam, S.; Lindahl, M.; Bullough, P.; Faruqi, A. R.; Tittor, J.; Oesterhelt, D.; Brown, L.; Lanyi, J. K. and Henderson, R. *J. Mol. Biol.*, 1999, 287:145–161.

[124] Miller, A. and Oesterhelt, D. *Biochim. Biophys. Acta*, 1990, 1020:57–64.

[125] Dioumaev, A. K.; Brown, L. S.; Needleman, R. and Lanyi, J. K. *Biochemistry*, 1998, 37:9889–9893.

[126] Dioumaev, A. K.; Brown, L. S.; Needleman, R. and Lanyi, J. K. *Biochemistry*, 1999, 38:10070–10078.

[127] Henderson, R.; Baldwin, J. M.; Ceska, T. A.; Zemlin, F.; Beckmann, K. H. and Downing, K. H. *J. Mol. Biol.*, 1990, 213:899–929.

[128] Landau, E. M. and Rosenbusch, J. P. *Proc. Natl. Acad. Sci. USA*, 1996, 93:14532–14535.

[129] Heberle, J.; Büldt, G.; Koglin, E.; Rosenbusch, J. P. and Landau, E. M. *J. Mol. Biol.*, 1998, 281:587–592.

[130] Hope, H. *Acta Cryst. B*, 1988, 44:22–26.

[131] Pebay-Peyroula, E.; Neutze R. and Landau, E. M. *Biochim. Biophys. Acta*, 2000, 1460:119–132.

[132] Neutze, R.; Pebay-Peyroula, E.; Edman, K.; Royant, A.; Navarro, J. and Landau, E. M. *Biochim. Biophys. Acta*, 2002, 1565:144–167.

[133] Luecke, H.; Richter, H. T. and Lanyi, J. K. *Science*, 1998, 280:1934–1937.

[134] Royant, A.; Grizot, S.; Kahn, R.; Belrhali, H.; Fieschi, F.; Landau, E. M. and Pebay-Peyroula, E. *Acta. Cryst. D*, 2002, 58:784–791.

[135] Edman, K.; Royant, A.; Larsson, G.; Jacobson, F.; Taylor, T.; van der Spoel, D.; Landau, E. M.; Pebay-Peyroula, E. and Neutze, R. *J. Biol. Chem.*, 2004, 279:2147–2158.

[136] Nave, C. *Radiat. Phys. Chem.*, 1995, 45:483–490.

[137] Ravelli, R. B. G. and McSweeney, S. M. *Structure*, 2000, 8:315–328.

[138] Weik, M.; Ravelli, R. B. G.; Kryger, G.; McSweeney, S.; Raves, M. L.; Harel, M.; Gros, P.; Silman, I.; Kroon, J. and Sussman, J. L. *Proc. Natl. Acad. Sci. USA*, 2000, 97:623–628.

[139] Matsui, Y.; Sakai, K.; Murakami, M.; Shiro, Y.; Adachi, S.; Okumura, H. and Kouyama, T. *J. Mol. Biol.*, 2002, 324:469–481.

[140] Réat, V.; Patzelt, H.; Ferrand, M.; Pfister, C.; Oesterhelt, D. and Zaccai, G. *Proc. Natl. Acad. Sci. USA*, 1998, 95:4970–4975.

[141] Ormos, P. *Proc. Natl. Acad. Sci. USA*, 1991, 88:473–477.

[142] Hajdu, J. and Andersson, I. *Annu. Rev. Biophys. Biomol. Struct.*, 1993, 22:467–498.

[143] Schlichting, I.; Berendzen, J.; Phillips Jr., G. N. and Sweet, R. M. *Nature*, 1994, 371:808–812.

[144] Genick, U.; Borgstahl, G. E. O.; Ng, K.; Ren, Z.; Pradervand, C.; Burke, P. M.; Srajer, V.; Teng, T. Y.; Schildkamp, W.; McRee, D. E.; Moffat, K. and Getzoff, E. D. *Science*, 1997, 275:1471–1475.

[145] Genick, U.; Soltis, S. M.; Kuhn, P.; Canestrelli, I. L. and Getzoff, E. D. *Nature*, 1998, 392:206–209.

[146] Chu, K.; Vojtchovsky, J.; McMahon, B. H.; Sweet, R. M.; Berendzen, J. and Schlichting, I. *Nature*, 2000, 403:921–923.

[147] Stowell, M. H. B.; McPhillips, T. M. ; Rees, D. C.; Soltis, S. M.; Abresch, E. and Feher, G. *Science*, 1997, 276:812–816.

[148] Srajer, V.; Teng, T. Y.; Ursby, T.; Pradervand, C.; Ren, Z.; Adachi, S. I.; Schildkamp, W.; Bourgeois, D.; Wulff, M.; and Moffat, K. *Science*, 1996, 274:1726–1729.

[149] Perman, B.; Srajer, V.; Ren, Z.; Teng, T. Y.; Pradervand, C.; Ursby, T.; Bourgeois, D.; Schotte, F.; Wulff, M.; Kort, R.; Hellingwerf, K. and Moffat, K. *Science*, 1998, 279:1946–1950.

[150] Lin, S. W. and Mathies, R. A. *Biophys. J.*, 1989, 56:653–660.

[151] Hirai, T. and Subramaniam, S. *FEBS Lett.*, 2003, 545:2–8.

[152] Gat, Y. and Sheves, M. *J. Am. Chem. Soc.*, 1993, 115:37723773.

[153] Tajkhorshid, E.; Baudry, J.; Schulten, K. and Suhai, S. *Biophy. J.*, 2000, 78:683–693.

[154] Lansing, J. C.; Hohwy, M.; Jaroniec, C. P.; Creemers, A. F. L.; Lugtenburg, J.; Herzfeld, J. and Griffin, R. G. *Biochemistry*, 2002, 41:431–438.

[155] Aharoni, A.; Khatchatouriants, A.; Manevitch, A.; Lewis, A. and Sheves, M. *J. Phys. Chem. B*, 2003, 107:6221 – 6225.

[156] Edman, K.; Nollert, P.; Royant, A.; Belrhali, H.; Pebay-Peyroula, E.; Hajdu, J.; Neutze, R. and Landau, E. M. *Nature*, 1999, 401:822–826.

[157] Schobert, B.; Cupp-Vickery, J.; Hornak, V.; Smith, S. O. and Lanyi, J. K. *J. Mol. Biol.*, 2002, 321:715–726.

[158] Bullough, P. A. and Henderson, R. *J. Mol. Biol.*, 1999, 286:1663–1671.

[159] Kandori, H.; Kinoshita, N.; Yamazaki, Y.; Maeda, A.; Shichida, Y.; Needleman, R.; Lanyi, J. K.; Bizounok, M.; Herzfeld, J.; Raap, J. and Lugtenburg, J. *Proc. Natl. Acad. Sci. USA*, 2000, 97:4649–4653.

[160] Roepe, P.; Gray, D.; Lugtenburg, J.; Van den Berg, E. M. M.; Herzfeld, J. and Rothschild, K. J. *J. Am. Chem. Soc.*, 1988, 110:7223 – 7224.

[161] Braiman, M. S.; Mogi, T.; Marti, T.; Stern, L. J.; Khorana, H. G. and Rothschild, K. J. *Biochemistry*, 1988, 27:8516–8520.

[162] Ahl, P. L.; Stern, L. J.; During, D.; Mogi, T.; Khorana, H. G.; and Rothschild, K. J. *J. Biol. Chem.*, 1988, 263:13594–13601.

[163] Kandori, H.; Belenky, M. and Herzfeld, J. *Biochemistry*, 2002, 41:6026–6031.

[164] Kandori, H.; Yamazaki, Y.; Shichida, Y.; Raap, J.; Lugtenburg, J.; Belenky, M. and Herzfeld, J. *Proc. Natl. Acad. Sci. USA*, 2001, 98:1571–1576.

[165] Sasaki, J.; Yuzawa, T.; Kandori, H.; Maeda, A. and Hamaguchi, H. *Biophys. J.*, 1995, 68:2073–2080.

[166] Maeda, A.J.; Sasaki, J.; Pfefferle, J.M.; Shichida, Y. and Yoshizawa, T. *Photochem. Photobiol.*, 1991, 54:911–921.

[167] Hage, W.; Kim, M.; Frei, H. and Mathies, R. A. *J. Phys. Chem.*, 1996, 100:16026–16033.

[168] Yamamoto, N.; Ebbesen, T. W. and Ohtani, H. *Chem. Phys. Lett.*, 1994, 228:61–65.

[169] Sasaki, J.; Maeda, A.; Kato, C. and Hamaguchi, H. *Biochemistry*, 1993, 32:867–871.

[170] Weidlich, O. and Siebert, F. *Appl. Spect.*, 1993, 47:1394.

[171] Royant, A.; Edman, K.; Ursby, T.; Pebay-Peyroula, E.; Landau, E. M. and Neutze, R. *Nature*, 2000, 406:645–648.

[172] Lanyi, J. K. and Schobert, B. *J. Mol. Biol.*, 2003, 328:439–450.

[173] Kouyama, T.; Nishikawa, T.; Tokuhisa, T. and Okumura, H. *J. Mol. Biol.*, 2004, 335:531–546.

[174] Lanyi, J. K. and Schobert, B. *J. Mol. Biol.*, 2007, 365:1379–1392.

[175] Hendrickson, F. M.; Burkard, F. and Glaeser, R. M. *Biophys. J.*, 1998, 75:1446–1454.

[176] Balashov, S. P. and Ebrey, T. G. *Photochem. Photobiol.*, 2001, 73:453462.

[177] Lanyi, J. K. and Luecke, H. *Curr. Opin. Struc. Biol.*, 2001, 11:415–419.

[178] Fodor, S. P. A.; Ames, J. B.; Gebhard, R.; Van den Berg, E. M. M.; Stoeckenius, W.; Lugtenburg, J. and Mathies, R. A. *Biochemistry*, 1988, 27:7097 – 7101.

[179] Mak-Jurkauskas, M. L.; Bajaj, V. S.; Hornstein, M. K.; Belenky, M.; Griffin, R. G.; Herzfeld, J. *Proc. Natl. Acad. Sci. USA*, 2008, 105:883–888.

[180] Morgan, J. E.; Gennis, R. B. and Maeda, A. *Photochem. Photobiol.*, 2008, 84:1038–1045.

[181] Sass, H. J.; Büldt, G.; Gessenich, R.; Hehn, D.; Neff, D.; Schlesinger, R.; Berendzen, J. and Ormos, P. *Nature*, 2000, 406:649–653.

[182] Luecke, H. *Biochim. Biophys. Acta*, 2000, 1460:133–151.

[183] Herzfeld, J. and Tounge, B. *Biochim. Biophys. Acta Bioenerg.*, 2000, 1460:95–105.

[184] Herzfeld, J. and Lansing, J. C. *Annu. Rev. Biophys. Biomol. Struct.*, 2002, 31:73–95.

[185] Lanyi, J. K. *J. Phys. Chem. B*, 2000, 104:11441–11448.

[186] Betancourt, F. M. H. and Glaeser, R. M. *Biochim. Biophys. Acta*, 2000, 1460:106–118.

[187] Facciotti, M. T.; Rouhani, S. and Glaeser, R. M. *FEBS Lett.*, 2004, 564:301–306.

[188] Lozier, R. H.; Xie, A.; Hofrichter, J. and Clore, G. M. *Proc. Natl. Acad. Sci. USA*, 1992, 89:3610–3614.

[189] Chizhov, I.; Chernavskii, D. S.; Engelhard, M.; Mueller, K. H.; Zubov, B. V.; Hess, B. *Biophys. J.*, 1996, 71:2329–2345.

[190] Váró, G. and Lanyi, J. K. *Biochemistry*, 1991, 30:5008–5015.

[191] Váró, G. and Lanyi, J. K. *Biochemistry*, 1991, 30:5016–5022.

[192] Váró, G. and Lanyi, J. K. *Biochemistry*, 1991, 30:7165–7171.

[193] Lanyi, J. K. *Isr. I. Chem.*, 1995, 35:365–385.

[194] Ames, J. B. and Mathies, R. A. *Biochemistry*, 1990, 29:7181–7190.

[195] Nagle, J. F. *Photochem. Photobiol.*, 1991, 54:897–903.

[196] Sass, H. J.; Schachowa, I. W.; Rapp, G.; Koch, M. H. J.; Oesterhelt, D.; Dencher, N. A. and Büldt, G. *EMBO J.*, 1997, 16:14841491.

[197] Sass, H.J.; Gessenich, R.; Koch, M.H.J.; Oesterhelt, D.; Dencher, N.A.; Büldt, G. and Rapp, G. *Biophys. J.*, 1998, 75:399–405.

[198] Rödig, C.; Chizhov, I.; Weidlich, O. and Siebert, F. *Biophys. J.*, 1999, 76:2687–2701.

[199] Dencher, N. A.; Dresselhaus, D.; Zaccai, G. and Büldt, G. *Proc. Natl. Acad. Sci. USA*, 1989, 86:7876–7879.

[200] Nakasako, M.; Kataoka, M.; Amemiya, Y. and Tokunaga, F. *FEBS Lett.*, 1991, 292:73–75.

[201] Han, B. G.; Vonck, J. and Glaeser, R. M. *Biophys. J.*, 1994, 67:1179–1186.

[202] Vonck, J. *Biochemistry*, 1996, 35:5870–5878.

[203] Weik, M.; Zaccai, G.; Dencher, N. A.; Oesterhelt, D. and Hauss, T. *J. Mol. Biol.*, 1998, 275:625–634.

[204] Zimányi, L.; Váró, G.; Chang, M.; Ni, B.; Needleman, R. and Lanyi, J. K. *Biochemistry*, 1992, 31:8535 – 8543.

[205] Kataoka, M. and Kamikubo, H. *Biochim. Biophys. Acta*, 2000, 1460:166–176.

[206] Lanyi, J. K. and Schobert, B. *J. Mol. Biol.*, 2002, 321:727–737.

[207] Facciotti, M. T.; Rouhani, S.; Burkard, F. T.; Betancourt, F. M.; Downing, K. H.; Rose, R. B.; McDermott, G. and Glaeser, R. M. *Biophhys. J.*, 2001, 81:3442–3455.

[208] Schobert, B.; Brown, L. S. and Lanyi, J. K. *J. Mol. Biol.*, 2003, 330:553–570.

[209] Cao, Y.; Brown, L. S.; Sasaki, J.; Maeda, A.; Needleman, R. and Lanyi, J. K. *Biophys. J.*, 1995, 68:1518–1530.

[210] Luecke, H.; Schobert, B.; Richter, H. T.; Cartailler, J. P. and Lanyi, J. K. *Science*, 1999, 286:255–261.

[211] Takeda, K.; Matsui, Y.; Kamiya, N.; Adachi, S.; Okumura, H. and Kouyama, T. *J. Mol. Biol.*, 2004, 341:1023–1037.

[212] Luecke, H.; Schobert, B.; Cartailler, J. P.; Richter, H. T.; Rosengarth, A.; Needleman, R. and Lanyi, J. K. *J. Mol. Biol.*, 2000, 300:1237–1255.

[213] Subramaniam, S. and Henderson, R. *Nature*, 2000, 406:653–657.

[214] Govindjee, R.; Misra, S.; Balashov, S. P.; Ebrey, T. G.; Crouch, R.K. and Menick, D. R. *Biophys. J.*, 1996, 71:1011–1023.

[215] Brown, L. S.; Sasaki, J.; Kandori, H.; Maeda, A.; Needleman, R. and Lanyi, J. K. *J. Biol. Chem.*, 1995, 270:27122 – 27126.

[216] Balashov, S.P.; Imasheva, E. S.; Ebrey, T. G.; Chen, N.; Menick, D. R. and Crouch, R. K. *Biochemistry*, 1997, 36:8671–8676.

[217] Kalaidzidis, I. V.; Belevich, I. N. and Kaulen, A. D. *FEBS Lett.*, 1998, 434:197–200.

[218] Essen, L.O.; Siegert, R.; Lehmann, W.D. and Oesterhelt, D. *Proc. Natl. Acad. Sci. USA*, 1998, 95:11673–11678.

[219] Zscherp, C.; Schlesinger, R.; Tittor, J.; Oesterhelt, D. and Heberle, J. *Proc. Natl. Acad. Sci. USA*, 1999, 96:5498–5503.

[220] Rammelsberg, R.; Huhn, G.; Lübben, M. and Gerwert, K. *Biochemistry*, 1998, 37:5001 – 5009.

[221] Spassov, V. Z.; Luecke, H.; Gerwert, K. and Bashford, D. *J. Mol. Biol.*, 2001, 312:203–219.

[222] Garczarek, F.; Brown, L.S.; Lanyi, J.K. and Gerwert, K. *Proc. Natl. Acad. Sci. USA*, 2005, 102:3633–3638.

[223] Garczarek, F. and Gerwert, K. *Nature*, 2006, 439:109–112.

[224] Xiao, Y.; Hutson, M. S.; Belenky, M.; Herzfeld, J. and Brainman, M. S. *Biochemistry*, 2004, 43:12809–12818.

[225] Balashov, S. P.; Imasheva, E. S.; Govindjee, R. and Ebrey, T. G. *Biophys. J.*, 1996, 70:473–481.

[226] Otto, H.; Marti, T.; Holz, M.; Mogi, T.; Stern, L. J.; Engel, F.; Khorana, H. G. and Heyn, M. P. *Proc. Natl. Acad. Sci. USA*, 1990, 87:1018–1022.

[227] Kalisky, O.; Ottolenghi, M.; Honig, B. and Korenstein, R. *Biochemistry*, 1981, 20:649–655.

[228] Tanford, C. *Proc. Natl. Acad. Sci. USA*, 1982, 79:2882–2884.

[229] Oesterhelt, D..; Tittor, J. and Bamberg, E. *J. Bioenerg. Biomembr.*, 1992, 24:181–191.

[230] Nagle, J. F. and Mille, M. *J. Chem. Phys*, 1981, 74:1367–1372.

[231] Schulten, K.; Schulten, Z. and Tavan, P. *Information and Energy Transduction in Biological Membranes*, chapter An isomerization model for the pump cycle of bacteriorhodopsin, pages 113–131. 1984, Alan R. Liss, New York.

[232] Tittor, J.; Paula, S.; Subramaniam, S.; Heberle, J.; Henderson, R. and Oesterhelt, D. *J. Mol. Biol.*, 2002, 319:555–565.

[233] Richter, H. T.; Needleman, R.; Kandori, H.; Maeda, A. and Lanyi, J. K. *Biochemistry*, 1996, 35:15461–15466.

[234] Tanimoto, T.; Furutani, Y. and Kandori, H. *Biochemistry*, 2003, 42:2300 – 2306.

[235] Gerwert, K.; Hess, B.; Soppa, J. and Oesterhelt, D. *Proc. Natl. Acad. Sci. USA*, 1989, 86:4943–4947.

[236] Otto, H.; Marti, T.; Holz, M.; Mogi, T.; Lindau, M.; Khorana, H. G. and Heyn, M. P. *Proc. Natl. Acad. Sci. USA*, 1989, 86:9228–9232.

[237] Bousché, O.; Braiman, M. S.; He, Y.-W.; Marti, T.; Khorana, H. G. and Rothschild, K. J. *J. Biol. Chem.*, 1991, 266:11063–11067.

[238] Braiman, M. S.; Bousche, O. and Rothschild, K. J. *Proc. Natl. Acad. Sci. USA*, 1991, 88:2388–2392.

[239] Souvignier, G. and Gerwert, K. *Biophys. J.*, 1992, 63:1393–1405.

[240] Holz, M.; Drachev, L. A.; Mogi, T.; Otto, H.; Kaulen, A. D.; Heyn, M. P.; Skulachev, V. P. and Khorana, H. G. *Proc. Natl. Acad. Sci. USA*, 1989, 86:2167–2171.

[241] Száraz, S.; Oesterhelt, D. and Ormos, P. *Biophy. J.*, 1994, 67:1706–1712.

[242] Brown, L. S. and Lanyi, J. K. *Proc. Natl. Acad. Sci. USA*, 1996, 93:1731–1734.

[243] Cao, Y.; Váró, G.; Klinger, A. L.; Czajkowsky, D. M.; Braiman, M. S.; Needleman, R. and Lanyi, J. K. *Biochemistry*, 1993, 32:1981–1990.

[244] Vonck, J. *EMBO J.*, 2000, 19:2152–2160.

[245] Wikström,M. *Curr. Opin. Struct. Biol.*, 1998, 8:480–488.

[246] Cao, Y.; Váró, G.; Chang, M.; Ni., B.; Needleman, R. and Lanyi, J. K. *Biochemistry*, 1991, 30:10972–10979.

[247] Váró, G. and Lanyi, J. K. *Biophy. J.*, 1991, 59:313–322.

[248] Váró, G. and Lanyi, J. K. *Biochemistry*, 1995, 34:12161–12169.

[249] Zhou, F.; Windemuth, A. and Schulten, K. *Biochemistry*, 1993, 32:2291–2306.

[250] Humphrey, W.; Logunov, I.; Schulten, K. and Sheves, M. *Biochemistry*, 1994, 33:3668–3678.

[251] Roux, B.; Nina, M.; Pomès, R. and Smith, J. C. *Biophy. J.*, 1996, 71:670–681.

[252] Murata, K.; Hoshino, T.; Sato, Y.; Hata, M. and Tsuda, M. *J. Mol. Struct. (Theochem)*, 2003, 664-665:125–133.

[253] Lee, Y. S. and Kraus, M. *J. Am. Chem. Soc.*, 2004, 126:2225–2230.

[254] Pfefferle, J. M.; Maeda, A.; Sasaki, J. and Yoshizawa, T. *Biochemistry*, 1991, 30:6548–6556.

[255] Martinez, L. C.; and Turner, G. J. *Biochim. Biophys. Acta*, 2002, 1564:91–98.

[256] Bousché, O.; Sonar, S.; Krebs, M. P.; Khorana, H. G. and Rothschild, K. J. *Photochem. Photobiol.*, 1992, 56:1085–1095.

[257] Balashov, S. P.; Lu, M.; Imasheva, E. S.; Govindjee, R.; Ebrey, T. G.; Othersen III, B.; Chen, Y.; Crouch, R. K. and Menick, D. R. *Biochemistry*, 1999, 38:2026–2039.

[258] Dioumaev, A. K.; Brown, L. S.; Needleman, R. and Lanyi, J. K. *Biochemistry*, 2001, 40:11308–11317.

[259] Balashov, S. P. *Biochim. Biophys. Acta*, 2000, 1460:75–94.

[260] Brown, L. S.; Yamazaki, Y.; Maeda, A.; Sun, L.; Needleman, R. and Lanyi, J. K. *J. Mol. Biol.*, 1994, 239:401–414.

[261] Brown, L. S.; Váró, G.; Needleman, R. and Lanyi, J. K. *Biophy. J.*, 1995, 69:2103–2111.

[262] Nachliel, E.; Yaniv-Checover, S. and Gutman, M. *Solid State Ionics*, 1997, 97:75–82.

[263] Checover, S.; Marantz, Y.; Nachliel, E.; Gutman, M.; Pfeiffer, M.; Tittor, J.; Oesterhelt, D. and Dencher, N. A. *Biochemistry*, 2001, 40:4281–4292.

[264] Riesle, J.; Oesterhelt, D.; Dencher, N. A. and Heberle, J. *Biochemistry*, 1996, 35:6635–6643.

[265] Friedman, R.; Nachliel, E. and Gutman, M. *Biophys. J.*, 2003, 85:886–896.

[266] Brown, L. S.; Needleman, R. and Lanyi, J. K. *Biochemistry*, 1999, 38:6855–6861.

[267] Schätzler, B.; Dencher, N. A.; Tittor, J.; Oesterhelt, D.; Yaniv-Checover, S.; Nachliel, E. and Gutman, M. *Biophy. J.*, 2003, 84:671–686.

[268] Coutre, J. L. and Gerwert, K. *FEBS Lett.*, 1996, 398:333–336.

[269] Delaney, J. K.; Yahalom, G.; Sheves, M. and Subramaniam, S. *Proc. Natl. Acad. Sci. USA*, 1997, 94:5028–5033.

[270] Subramaniam, S.; Greenhalgh, D. A.; Rath, P.; Rothschild, K. J. and Khorana, H. G. *Proc. Natl. Acad. Sci. USA*, 1991, 88:6873–6877.

[271] Delaney, J. K.; Schweiger, U. and Subramaniam, S. *Proc. Natl. Acad. Sci. USA*, 1995, 92:11120–11124.

[272] Smith, S. O.; Pardoen, J. A.; Mulder, P. P. J.; Curry, B.; Lugtenburg, J. and Mathies, R. *Biochemistry*, 1983, 22:6141 – 6148.

[273] Zscherp, C. and Heberle, J. *J. Phys. Chem. B*, 1997, 101:10542 – 10547.

[274] Rouhani, S.; Cartailler, J. P.; Facciotti, M. T.; Walian, P.; Needleman, R.; Lanyi, J. K.; Glaeser, R. M. and Luecke, H. *J. Mol. Biol.*, 2001, 313:615–628.

[275] Okumura, H.; Murakami, M. and Kouyama, T. *J. Mol. Biol.*, 2005, 351:481–495.

[276] Chen, D.; Wang, J. M. and Lanyi, J. K. *J. Mol. Biol.*, 2007, 366:790–805.

[277] Weidlich, O.; Schalt, B.; Friedman, N.; Sheves, M.; Lanyi, J. K.; Brown, L. S. and Siebert, F. *Biochemistry*, 1996, 35:10807–10814.

[278] Tavan, P.; Schulten, K. and Oesterhelt, D. *Biophys. J.*, 1985, 47:415–430.

[279] Tavan, P. and Schulten, K. *Biophy. J.*, 1986, 50:81–89.

[280] Smith, S. O.; Braiman, M. S.; Myers, A. B.; Pardoen, J. A.; Courtin, J. M. L.; Winkel, C.; Lugtenburg, J. and Mathies, R. A. *J. Am. Chem. Soc.*, 1987, 109:3108–3125.

[281] Smith, S. O.; Pardoen, J. A.; Lugtenburg, J. and Mathies, R. A. *J. Phys. Chem.*, 1987, 91:804–819.

[282] Fodor, S. P. A.; Pollard, W. T.; Gebhard, R.; Van Den Berg, E. M. M.; Lugtenburg, J and Mathies, R. A. *Proc. Natl. Acad. Sci. USA*, 1988, 85:2156–2160.

[283] Schulten, K. and Tavan, P. *Nature*, 1978, 272:85–86.

[284] Orlandi, G. and Schulten, K. *Chem. Phys. Lett.*, 1979, 64:370–374.

[285] Warshel, A. *Photochem. Photobiol.*, 1979, 30:285–290.

[286] Ben-Nun, M.; Molnar, F.; Lu, H.; Phillips, J. C.; Martínez, T.J. and Schulten, K. *Faraday Discuss.*, 1998, 110:447–462.

[287] Xu, D.; Martin, C. and Schulten, K. *Biophys. J.*, 1996, 70:453–460.

[288] Nonella, M.; Windemuth, A. and Schulten, K. *Photochem. Photobiol.*, 1991, 54:937–948.

[289] Nina, M.; Roux, B. and Smith, J. C. *Biophys. J.*, 1995, 68:25–39.

[290] Scheiner, S. and Hillenbrand, E. A. *Proc. Natl. Acad. Sci. USA*, 1985, 82:2741–2745.

[291] Scheiner, S. and Duan, X. *Biophys. J.*, 1991, 60:874–883.

[292] Tajkhorshid, E. and Suhai, S. *Chem. Phys. Lett.*, 1999, 5:457–464.

[293] Tajkhorshid, E.; Paizs, B. and Suhai, S. *J. Phys. Chem. B*, 1999, 103:45184527.

[294] Murata, K.; Fujii, Y.; Enomoto, N.; Hata, M.; Hoshino, T. and Tsuda, M. *Biophys. J.*, 2000, 79:982–991.

[295] Nakajima, S.; Ohno, K.; Inoue, Y. and Sakurai, M. *J. Phys. Chem. B*, 2003, 107:2867–2874.

[296] Song, Y.; Mao, J. and Gunner, M. R. *Biochemistry*, 2003, 42:9875–9888.

[297] Hayashi, S. and Ohmine, I. *J. Phys. Chem.*, 2000, 104:10678–10691.

[298] Bondar, A. -N.; Suhái, S.; Fischer, S.; Smith, J. C. and Elstner, M. *J. Struct. Biol.*, 2007, 157:454–469.

[299] Bondar, A.-N.; Baudry, J.; Suhai, S.; Fischer,S. and Smith, J. C. *J. Phys. Chem. B*, 2008, 112:1472914741.

[300] Bondar, A. -N.; Fischer, S.; Smith, J. C.; Elstner, M. and Suhái, S. *J. Am. Chem. Soc.*, 2004, 126:14668–14677.

[301] Braun-sand, S.; Sharma, P.; Chu, Z.; Pisliakov, A. V. and Warshel, A. *Biochim. Biophy. Acta*, 2008, 1777:441–452.

[302] Warshel, A. *Acc. Chem. Res.*, 1981, 14:284–290.

[303] Rousseau, R.; Kleinschmidt, V.; Schmitt, U. W. and Marx, D. *Phys. Chem. Chem. Phys.*, 2004, 6:1848–1859.

[304] Mathias, G. and Marx, D. *Proc. Natl. Acad. Sci. USA*, 2007, 104:6980–6985.

[305] Xu, D.; Sheves, M. and Schulten, K. *Biophys. J.*, 1995, 69:2745–2760.

[306] Hermone, A. and Kuczera, K. *Biochemistry*, 1998, 37:2843–2853.

[307] Watanabe, H. C.; Ishikura, T. and Yamato, T. *Proteins*, Apr 2009, 75:53–61.

[308] Szabo, A. and Ostlund, N. S. *Modern Quantum Chemistry*. 1982, Macmillan, New York.

[309] Blinder, S. M. *Am. J. Phys.*, 1965, 33:431–443.

[310] Cramer, C. *Essentials of Computational Chemistry: Theories and Models*. Second Edition, John Wiley & Sons Ltd.

[311] Slater, J. C. *Phys. Rev.*, 1930, 36:57–64.

[312] Boys, S. F. *Proc. R. Soc. London Ser. A*, 1950, 200:542–554.

[313] Hehre, W. J.; Stewart, R. F. and Pople, J. A. *J. Chem. Phys.*, 1969, 51:2657–2664.

[314] Woon, D. E. and Dunning, T. H. *J. Chem. Phys.*, 1993, 98:1358–1371.

[315] Woon, D. E. and Dunning, T. H. *J. Chem. Phys.*, 1995, 103:4572–4585.

[316] Levine, I. *Quantum Chemistry*. 1991, Prentice Hall Inc.

[317] Jensen, F. *Introduction to Computational Chemistry*. 1999, Wiley:Chichester.

[318] Borden, W. T. and Davidson, E. R. *Acc. Chem. Res.*, 1996, 29:67–75.

[319] Löwdin, P. O. *Int. J. Quant. Chem.*, 1995, 55:77–102.

[320] Neese, F. *J. Chem. Phys.*, 2003, 119(18):9428–9443.

[321] Wanko, M.; Hoffmann, M.; Strodel, P.; Koslowski, A.; Thiel, W.; Neese, F.; Frauenheim, T. and Elstner, M. *J. Phys. Chem. B*, 2005, 109(8):3606–3615.

[322] Schäfer, A.; Horn, H. and Ahlrichs, R. *J. Chem. Phys.*, 1992, 97:2571–2577.

[323] Dewar, M. J. S. *The Molecular Orbital Theory of Organic Chemistry*. 1969, McGraw-Hill, New York.

[324] Pople, J. A. and Beveridge, D. L. *Approximate Molecular Orbital Theory*. 1970, Academic Press, New York.

[325] Murrell, J. N. and Harget, A. J. *Semiempirical Self-Consistent-Field Molecular Orbital Theory of Molecules*. 1972, Wiley, New York.

[326] Segal, G. A. *Modern Theoretical Chemistry*. 1977, Plenum, New York.

[327] Clark, T. *A Handbook of Computational Chemistry*. 1985, Wiley, New York.

[328] Jug, K. *Theor. Chim. Acta*, 1980, 54:263–300.

[329] Dewar, M. J. S. *J. Phys. Chem.*, 1985, 89:2145–2150.

[330] Thiel, W. *Adv. Chem. Phys.*, 1996, 44:703–757.

[331] Thiel, W. *Modern Methods and Algorithms of Quantum Chemistry*, chapter Semiempirical Methods, pages 233–255. 2000, NIC Series, Juelich.

[332] Pople, J. A.; Santry, D. P. and Segal, G. A. *J. Chem. Phys.*, 1965, 43:S129–S135.

[333] Pople, J. A. and Segal, G. A. *J. Chem. Phys.*, 1965, 43:S136–S151.

[334] Pople, J. A.; Beveridge, D. L. and Dobosh, P. A. *J. Chem. Phys.*, 1967, 47:2026–2033.

[335] Ridley, J. and Zerner, M. C. *Theor. Chim. Acta*, 1973, 32:111–134.

[336] Bacon, A. D. and Zerner, M. C. *Theor. Chim. Acta*, 1979, 53:21–54.

[337] Dewar, M. J. S. and Thiel, W. *J. Am. Chem. Soc.*, 1977, 99:4899–4907.

[338] Dewar, M. J. S. and Thiel, W. *J. Am. Chem. Soc.*, 1977, 99:4907–4917.

[339] Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F. and Stewart, J. J. P. *J. Am. Chem. Soc.*, 1085, 107:3902–3909.

[340] Stewart, J.J. P. *J. Comp. Chem.*, 1989, 10:221–264.

[341] Parr, R. G. and Yang, W. *Density-functional theory of atoms and molecules*. 1989, Oxford Univ. Press.

[342] Hohenberg, P. and Kohn, W. *Phys. Rev.*, Nov 1964, 136:B864–B871.

[343] Jones, R. O. and Gunnarsson, O. *Rev. Mod. Phys.*, Jul 1989, 61:689–746.

[344] Kohn, W. and Sham, L. J. *Phys. Rev.*, Nov 1965, 140:A1133–A1138.

[345] Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, Th.; Suhái. S. and Seifert, G. *Phys. Rev. B*, 1998, 58:7260–7268.

[346] Porezag, D. ; Frauenheim, Th.; Köhler, Th.; Seifert, G. and Kaschner, R. *Phys. Rev. B*, 1995, 51:12947–12957.

[347] Foulkes, W.; Matthew, C. and Haydock, R. *Phys. Rev. B*, 1989, 39:12520–12536.

[348] Elstner, M. *Theor. Chem. Acc.*, 2006, 116:316–325.

[349] Yang, Y.; Yu, H.; York, D.; Cui, Q. and Elstner, M. *J. Phys. Chem. A*, 2007, 111:10861–10873.

[350] Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard III, W. A. and Skiff, W. M. *J. Am. Chem. Soc.*, 1991, 114:10024–10035.

[351] Allinger, N. L. *J. Am. Chem. Soc.*, 1977, 99:8127–8134.

[352] Allinger, N. L.; Yuh, Y. H. and Lii, J. H. *J. Am. Chem. Soc.*, 1989, 111:8551–8566.

[353] Lii, J. H. and Allinger, N. L. *J. Am. Chem. Soc.*, 1989, 111:8566–8575.

[354] Allinger, N. L .; Li, F. and Yan, L. *J. Comp. Chem.*, 1990, 11:848–867.

[355] Allinger, N. L.; Chen, K. and Lii, J. H. *J. Comp. Chem.*, 1996, 17:642–668.

[356] Allinger, N. L.; Chen, K.; Katzenelenbogen, J. A.; Wilson, S.R. and Anstead, G. M. *J. Comp. Chem.*, 1996, 17:747–755.

[357] Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S. and Karplus, M. *J. Comp. Chem.*, 1983, 4:187–217.

[358] Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S. and Weiner, P. *J. Am. Chem. Soc.*, 1984, 106:765–784.

[359] Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W. and Kollman, P. A. *J. Am. Chem. Soc.*, 1995, 117:5179–5197.

[360] Jorgensen, W. L.; Maxwell, D. S. and Tirado-Rives, J. *J. Am. Chem. Soc.*, 1996, 118:11225–11236.

[361] van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hnenberger, P. H.; Krger, P.; Mark, A. E.; Scott, W. R. P. and Tironi, I. G. *Biomolecular Simulation: The GROMOS96 Manual and User Guide*. 1996, ETH Zrich and BIOMOS Groningen,.

[362] Reiher, W. H. III. *Theoretical studies of hydrogen bonding*. PhD thesis, 1985, Harvard University.

[363] MacKerell, A. D. Jr. *J. Phys. Chem. B*, 1998, 102:3586–3616.

[364] Morse, P. M. *Phys. Rev.*, 1929, 34:57–64.

[365] London, F. *Zeitschrift für Physik*, 1930, 63:245–279.

[366] Leach, A. *Molecular modelling principles and applications*. 2001, Pearson Prentice hall.

[367] Warshel, A. and Levitt, M. *J. Mol. Biol.*, 1976, 103:227–249.

[368] Singh, U. C. and Kollman, P. A. *J. Comp. Chem.*, 1986, 7:718–730.

[369] Gao, J. *Acc. Chem. Res.*, 1996, 29:298–305.

[370] Senn, H. M. and Thiel, W. *Top. Curr. Chem.*, 2007, 268:173–290.

[371] Gao, J. and Truhlar, D.H. *Annu. Rev. Phys. Chem.*, 2002, 53:467–505.

[372] Warshel, A. *Annu. Rev. Biophys. Biomol. Struct.*, 2003, 32:425–443.

[373] Cui, Q. and Karplus, M. *Adv. Prot. Chem.*, 2003, 66:315–372.

[374] König, P. H.; Hoffmann, M.; Frauenheim, Th. and Cui, Q. *J. Phys. Chem. B*, 2005, 109:9082–9095.

[375] Zhang, Y.; Lee, T. S. and Yang, W. *J. Chem. Phys.*, 1999, 110:46–54.

[376] Antes, I. and Thiel, W. *J. Phys. Chem. A*, 1999, 103:92909295.

[377] Alary, F.; Poteau, R.; Heully, J-L.; Barthelat, J-C. and Daudey, J-P. *Theo. Chem. Acc.*, 2000, 104:174–178.

[378] DiLabio, G. A.; Hurley, M. M. and Christiansen, P. A. *J. Chem. Phys.*, 2002, 116(22):9578–9584.

[379] Théry, V.; Rinaldi, D.; Rivail, J-L.; Maigret, B. and Ferenczy, G. G. *J. Comput. Chem.*, 1994, 15:269–282.

[380] Monard, G.; Loos, M.; Théry, V.; Baka, K. and Rivail, J-L. *Int. J. Quant. Chem.*, 1996, 58:153–159.

[381] Philipp, D. M. and Friesner, R.A . *J. Comput. Chem.*, 1999, 20:1468–1494.

[382] Philipp, D. M. and Friesner, R. A. *J. Comput. Chem.*, 2000, 21:1442–1457.

[383] Gao, J.; Amara, P.; Alhambra, C. and Field, M. J. *J. Phys. Chem. A*, 1998, 102(24):4714–4721.

[384] Amara, P.; Field, M. J.; Alhambra, C. and Gao, J. *Theo. Chem. Acc.*, 2000, 104:336–343.

[385] Maseras, F. and Morokuma, K. *J. Comp. Chem.*, 1995, 16:1170–1179.

[386] Field, M. J.; Bash, P. A. and Karplus, M. *J. Comput. Chem.*, 1990, 11:700–733.

[387] Riccardi, D.; Li, G. and Cui, Q. *J. Phys. Chem. B*, 2004, 108:64676478.

[388] Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S. and Morokuma, K. *J. Phys. Chem.*, 1996, 100:1935719363.

[389] Sauer, J. and Sierka, M. *J. Comput. Chem.*, 2000, 21:1470–1493.

[390] Sherwood, P.; de Vries, A. H.; Collins, S. J.; Greatbanks, S. P.; Burton, N. A.; Vincent, M. A. and Hillier, I. H. *Faraday Discuss.*, 1997, 106:79 – 92.

[391] Rick, S. W. and Stuart, S. J. *Reviews in Computational Chemistry*, chapter Potentials and algorithms for incorporating polarizability in computer simulations, pages 89–146. 2002, Wiley, Hoboken, NJ.

[392] Ponder, J. W. and Case, D. A. *Adv. Prot. Chem.*, 2003, 66:27–85.

[393] Yu, H. and van Gunsteren, W. F. *Comput. Phys. Commun.*, 2005, 172:69–85.

[394] Thole, B. T. *Chem. Phys.*, 1981, 59:341–350.

[395] Lamoureux, G. and Roux, B. *J. Chem. Phys.*, 2003, 119:3025–3039.

[396] Rappe, A. K. and Goddard III, W. A. *J. Phys. Chem.*, 1991, 95:3358–3363.

[397] Wanko, M.; Hoffmann, M.; Fraehmcke, J.; Frauenheim, T. and Elstner, M. *J. Phys. Chem B*, 2008, 112:11468–11478.

[398] Cui, Q.; Elstner, M.; Kaxiras, E.; Frauenheim, Th. and Karplus, M. *J. Phys. Chem. B*, 2001, 105(2):569–585.

[399] Koehl, P. *Curr. Opin. Struct. Biol.*, 2006, 16:142–151.

[400] Sagui, C. and Darden, T. A. *Annu. Rev. Biophys. Biomol. Struct.*, 1999, 28:155–179.

[401] Kuwajima, S. and Warshel, A. *J. Chem. Phys.*, 1988, 89:3751–3759.

[402] Cramer, C. J. and Truhlar, D. G. *Reviews in Computational Chemistry*, chapter Continuum Solvation Models: Classical and Quantum Mechanical Implementations, pages 1–72. 1995, VCH Publishers, New York.

[403] Nam, K.; Gao, J. and York, D. M. *J. Chem. Theo. Comp.*, 2005, 1:2–13.

[404] Weber, W.; Hünenberger, P. H. and, McCammon, J. A. *J. Phys. Chem. B*, 2000, 104:3668–3675.

[405] Schaefer, P.; Riccardi, D. and Cui, Q. *J. Chem. Phys.*, 2005, 123:14905–14914.

[406] Honig, B. and Nicholls, A. *Science*, 1995, 268:1144–1149.

[407] Warshel, A. and Papazyan, A. *Curr. Opin. Struct. Biol.*, 1998, 8:211–217.

[408] Fersht, A. *Enzyme Structure, Mechanism and Protein Foldings*. 1998, Freeman, San Francisco.

[409] Steinbach, P. J. and Brooks, B. R. *J. Comput. Chem.*, 1994, 15:667–683.

[410] Stote, R. H. and Karplus, M. *Proteins*, 1995, 23:12–31.

[411] Feller, S. E.; Pastor, R. W.; Rojnuckarin, A.; Bogusz, S. and Brooks, B. R. *J. Phys. Chem.*, 1996, 100:17011–17020.

[412] Buncel, E.; Stairs, R. A. and Wilson, H. *The Role of the Solvent in Chemical Reactions*. 2003, Oxford University Press.

[413] Sharp, K. A. and Honig, B. *Annu. Rev. Biophys.*, 1990, 19:301–332.

[414] Roux, B. and Simonson, T. *Biophys. Chem.*, 1999, 78:1–20.

[415] Cramer, C. J. and Truhlar, D.G. *Chem. Revs.*, 1999, 99:21612200.

[416] Feig, M. and Brooks, C. L. III. *Curr. Opin. Struct. Biol.*, 2004, 14:217–224.

[417] Lu, B. Z.; Chen, W. Z.; Wang, C. X. and Xu, X. J. *Proteins*, 2002, 48:497–504.

[418] Prabhu, N. V.; Zhu, P. and Sharp, K. A. *J. Comput. Chem.*, 2004, 25:2049–2064.

[419] Bashford, D. and Case, D. A. *Annu. Rev. Phys. Chem.*, 2000, 51:129–152.

[420] Feig, M.; Onufriev, A.; Lee, M.S.; Im, W.; Case, D.A. and Brooks, C.L. III. *J. Comput. Chem.*, 2004, 25:265–284.

[421] Lee, B. and Richards, F. M. *J. Mol. Biol.*, 1971, 55:379–380.

[422] Hermann, R. B. *J. Phys. Chem.*, 1972, 76:27542759.

[423] Dinner, A. R.; Lopez, X. and Karplus, M. *Theo. Chem. Acc.*, 2003, 109:118–124.

[424] Simonson, T.; Archontis, G. and Karplus, M. *J. Phys. Chem. B*, 1997, 101:8349–8362.

[425] Beglov, D. and Roux, B. *J. Chem. Phys.*, 1994, 100:9050–9063.

[426] Im, W.; Berneche, S. and Roux, B. *J. Chem. Phys.*, 2001, 114:2924–2937.

[427] Gao. J. *Reviews in Computational Chemistry*, chapter Continuum Solvation Models: Classical and Quantum Mechanical Implementations, pages 119–185. 1996, VCH Publishers, New York.

[428] Heberle, J. and Dencher, N. A. *FEBS Lett.*, 1990, 277:277–280.

[429] Cao, Y.; Brown, L.S.; Needleman, R. and Lanyi, J.K. *Biochemistry*, 1993, 32:10239–10248.

[430] Heberle, J. and Dencher, N. A. *Proc. Natl. Acad. Sci. USA*, 1992, 89:5996–6000.

[431] Headrick, J.M.; Diken, E. G.; Walters, R. S.; Hammer, N. I.; Christie, R. A.; Cui, J.; Myshakin, E.M.; Duncan, M.A.; Johnson, M.A. and Jordan, K. D. *Science*, 2005, 308:1765–1769.

[432] Ball, P. *Chem. Rev.*, 2008, 108:74–108.

[433] Harris, T.K. and Mildvan, A. S. *Proteins: Struct. Funct. and Gene.*, 1999, 35:275–282.

[434] Phatak, P.; Ghosh, N.; Yu, H.; Cui, Q. and Elstner, M. *Proc. Natl. Acad. Sci. USA*, 2008, 105:19672–19677.

[435] Brooks, B.R.; Bruccoleri, R.E.; Olafson, B.D.; States, D.J.; Swaminathan, S. and Karplus, M. *Journal of Computational Chemistry*, 1983, 4(2):187–217.

[436] Brunger, A. T. and Karplus, M. *Proteins*, 1988, 4:148–156.

[437] MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D. and Karplus, M. *J. Phys. Chem. B*, 1998, 102:3586–3616.

[438] Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W. and Klein, M. L. *J. Chem. Phys.*, 1983, 79:926–935.

[439] Tajkhoshid, E. and Suhai, S. *J. Phys. Chem. B*, 1999, 103:5581–5590.

[440] Tajkhoshid, E.; Paizs, B. and Suhai, S. *J. Phys. Chem. B*, 1999, 103:4518–4527.

[441] Im, W.; Berneche, S. and Roux, B. *J. Chem. Phys.*, 2001, 114:2924–2937.

[442] Brooks III, C. L. and Karplus, M. *J. Mol. Biol.*, 1989, 208:159–181.

[443] Steinbach, P. J. and Brooks, B. R. *J. Comput. Chem.*, 1994, 15:667–683.

[444] Nina, M.; Beglov, D. and Roux, B. *J. Phys. Chem. B*, 1997, 101:5239–5248.

[445] Nina, M.; Im, W. and Roux, B. *Biophys. Chem.*, 1999, 78:89–96.

[446] Woo, H. J.; Dinner, A. R. and Roux, B. *J. Chem. Phys.*, 2004, 121:6392–6400.

[447] Yang, Y.; Yu, H.; York, D.; Cui, Q. and Elstner, M. *J. Phys. Chem. A*, 2007, 111:10861–10873.

[448] Elstner, M.; Cui, Q.; Munih, P.; Kaxiras, E.; Frauenheim, T. and Karplus, M. *J. Comp. Chem.*, 2003, 24:565.

[449] Riccardi, D.; Schaefer, P.; Yang, Y.; Yu, H.; Ghosh, N.; Prat-Resina, X.; Koenig, P.; Li, G.; Xu, D.; Guo, H.; Elstner, M. and Cui, Q. *J. Phys. Chem. B*, 2006, 110:6458–6469.

[450] Hoffmann, M.; Wanko, M.; Strodel, P.; Koenig, P.; Frauenheim, T.; Schulten, K.; Thiel, W.; Tajkhorshid, E. and Elstner, M. *J. Am. Chem. Soc.*, 2006, 128:10808–10818.

[451] Riccardi, D.; Schaefer, P. and Cui, Q. *J. Phys. Chem. B*, 2005, 109:17715.

[452] Riccardi, D. and Cui, Q. *J. Phys. Chem. A*, 2007, 111:5703–5711.

[453] Gordon, R. G. *J. Chem. Phys.*, 1965, 43:1307.

[454] McQuarrie, D. A. *Statistical Mechanics*. Harper and Row, 1976, New York.

[455] Yu, H. and Cui, Q. *J. Chem. Phys.*, 2007, 127:234504.

[456] Dioumaev, A. K. and Braiman, M. S. *J. Am. Chem. Soc.*, 1995, 117:10572–10574.

[457] Kandt, C.; Schlitter, J. and Gerwert, K. *Biophys. J.*, 2004, 86:705–717.

[458] Roscioli, J. R.; McCunn, L. R. and Johnson, M. A. *Science*, 2007, 316:249–254.

[459] Perrin, C. L. and Nielson, J. B. *Annu. Rev. Phys. Chem.*, 1997, 48:511–544.

[460] Cleland, W. W.; Ferry, P. A. and Gerlt, J. A. *J. Biol. Chem.*, 1998, 273:25529–25532.

[461] Balashov, S. P.; Govindjee, R.; Imasheva, E. S.; Misra, S.; Ebrey, T. G.; Feng. Y.; Crouch, R. K. and Menick, D. R. *Biochemistry*, 1995, 34:8820–8834.

[462] Chizhov, I.; Engelhard, M.; Chernavskii, D. S.; Zubov, B. and Hess, B. *Biophys. J.*, 1992, 61:1001–1006.

[463] Zscherp, C.; Schlesinger, R. and Heberle, J. *Biochem. Biophys. Res. Comm.*, 2001, 283:57–63.

[464] Efremov, R.; Gordeliy, V. I.; Heberle, J. and Büldt, G. *Biophy. J.*, 2006, 91:1441–1451.

[465] Zhang, X.; Harrison, D. V. T. and Cui, Q. *J. Am. Chem. Soc.*, 2002, 124:14871–14878.

[466] Cui, Q.; Elstner, M. and Karplus, M. *J. Phys. Chem. B*, 2002, 106:2721–2740.

[467] Li, G. and Cui, Q. *J. Am. Chem. Soc.*, 2003, 125:15028–15038.

[468] Li, G.; Zhang, X. and Cui, Q. *J. Phys. Chem. B*, 2003, 107:8643–8653.

[469] Elstner, M.; Frauenheim, T. and Suhái, S. *J. Mol. Struct:THEOCHEM*, 2003, 632:29–41.

[470] Riccardi, D.; König, P.; Prat-Resina, X.; Yu, H.; Elstner, M.; Frauenheim, T. and Cui, Q. *J. Am. Chem. Soc.*, 2006, 128:16302–16311.

[471] Zhou, H.; Tajkhorshid, E.; Frauenheim, T.; Suhái, S. and Elstner, M. *Chem. Phys.*, 2002, 277:91–103.

[472] Bondar, A.-N.; Fischer, S.; Suhai, S. and Smith, J. C. *J. Phys. Chem. B*, 2005, 109:14786–14788.

[473] Okada, T.; Sugihara, M.; Bondar, A.-N.; Elstner, M.; Entel, P. and Buss, V. *J. Mol. Biol.*, 2004, 342:571–583.

[474] Wanko, M.; Hoffmann, M; Frauenheim, T. and Elstner, M. *J. Phys. Chem B*, 2008, 112:11462–11467.

[475] Warshel, A. and Chu, Z.T. *J. Phy. Chem. B*, 2001, 105:9857–9871.

[476] Hayashi, S.; Tajkhorshid, E.; Schulten, K. *Biophys. J.*, 2003, 85:1440–1449.

[477] Hayashi, S.; Tajkhorshid, E. and Schulten, K. *Biophys. J.*, 2002, 83:1281–1297.

[478] Hayashi, S.; Tajkhorshid, E. and Schulten, K. *J. Am. Chem. Soc.*, 2004, 126:10516 – 10517.

[479] Ren, L.; Martin, C. H.; Wise, K. J.; Gillespie, N. B.; Luecke, H.; Lanyi, J. K.; Spudich, J. L. and Birge, R. R. *Biochemistry*, 2001, 40(46):13906–13914.

[480] Hayashi, S.; Tajkhorshid, E.; Pebay-Peyroula, E.; Royant, R.; Landau, E.; Navarro, J. and Schulten, K. *J. Phys. Chem B*, 2001, 105:10124–10131.

[481] Fujimoto, K.; Hasegawa, J.; Hayashi, S.; Kato, S. and Nakatsuji, H. *Chem. Phys. Lett.*, 2005, 414:239–242.

[482] Fujimoto, K.; Hayashi, S.; Hasegawa, J. and Nakatsuji, H. *J. Chem. Th. and Comp.*, 2007, 3:605–618.

[483] Matsuura, A.; Sato, H.; Houjou, H.; Saito, S.; Hayashi, T. and Sakurai, M. *J. Comp. Chem.*, 2006, 27:1623–1630.

[484] Ghosh, N. and Cui, Q. *J. Phys. Chem.*, 2008, 112:8387–9397.

[485] Elstner, M. and Cui, Q. *Challenges and Advances in Computational Chemistry and Physics, Vol.6*, chapter Combined QM/MM methods for the simulation of condensed phase processes using an approximate DFT approach, pages 381–405. 2008, Springer.

[486] Cui, Q. and Elstner, M. *Multi-scale Quantum Models for Biocatalysis: Modern Techniques and Applications*, chapter Multi-scale QM/MM methods with Self-Consistent-Charge Density-Functional-Tight-Binding (SCC-DFTB), page In Press. , Springer.

[487] Fischer, S. and Karplus, M. *Chem. Phys. Lett.*, 1992, 194:252–261.

[488] Choi, C. and Elber, R. *J. Chem. Phys.*, 1991, 94(1):751–760.

[489] Neese, F. *ORCA - An ab initio, density functional and semiempirical program package*. 2004, Version 2.3 - Revision 09; Max Planck Institut fuer Strahlenchemie: Muelheim, Germany.

[490] Cui, Q. and Karplus, M. *J. Chem. Phys.*, 2000, 112(3):1133–1149.

[491] Yu, H. and Cui, Q. *J. Chem. Phys.*, 2007, 127:234504.

[492] Brooks III, C. L. and Karplus, M. *J. Mol. Biol.*, 1989, 208:159–181.

[493] Rychaert, J. P.; Ciccotti, G. and Berendsen, H. J. *J. Comput. Phys.*, 1977, 23:327–341.

[494] Hessling, B; Souvignier, G. and Gerwert, K. *Biophy. J.*, 1993, 65:1929–1941.

[495] Simmons, C. J.; Liu, R. S. H.; Denny, M. and Seff, K. *Acta Crystallogr. B*, 1981, 37:2197–2205.

[496] Hamanaka, T.; Mitsui, T.; Ashida, T. and Kakudo, M. *Acta Crystallogr. B*, 1972, 28:214–222.

[497] Santarsiero, B. D.; James, M. N. G.; Mahendran, M. and Childs, R. F. *J. Am. Chem. Soc.*, 1990, 112:9416–9418.

[498] El-Sayed, M. A.; Lin, C. T. and Mason, W. R. *Proc. Natl. Acad. Sci. USA*, 1989, 86:5376–5379.

[499] Wu, S. and El-Sayed, M. A. *Biophy. J.*, 1991, 60:190–197.

[500] Smith, S.O.; Lugtenburg, J. and Mathies, R. *J. Memb. Biol.*, 1985, 85:95–109.

[501] Subramaniam, S.; Faruqi, A. R.; Oesterhelt, D. and Henderson, R. *Proc. Natl. Acad. Sci. USA*, 1997, 94:1767–1772.

[502] Lanyi, J. K. *Biochim. Biophys. Acta*, 2006, 1757:1012–1018.